HARWELL
Research Laboratory of the UKAEA

25 years of theoretical physics 1954 — 1979

A special progress report from
Theoretical Physics Division, Harwell
in recognition of the 25th anniversary of the
U.K. Atomic Energy Authority

UNCLASSIFIED TP.843

25 Years of Theoretical Physics 1954 - 1979

A special progress report from Theoretical Physics Division, Harwell in recognition of the 25th anniversary of the U.K. Atomic Energy Authority

Theoretical Physics Division,
A.E.R.E., Harwell, Oxfordshire.
March, 1980.

CONTENTS

List of Contributors

Dr. J.S. Bell, European Organization for Nuclear Research, CERN, FH-1211 Geneva 23, Switzerland.

Dr. J.S. Briggs, c/o Fakultät für Physik, der Universität Freiburg IBR, 78 Freiburg IBR, Hermann-Herder-Strasse 3, West Germany.

Dr. R. Bullough, Theoretical Physics Division, Building 424.4, A.E.R.E., Harwell, Oxon, OX11 ORA.

Prof. P.G. Burke, Daresbury Laboratory, Science Research Council, Daresbury, Warrington, WA4 4AD.

Dr. Ian Cheshire, C.S.S.D., Building 6, A.E.R.E, Harwell, Oxon, OX11 ORA.

Mr. A.R. Curtis, C.S.S.D., Building 8.9, A.E.R.E., Harwell, Oxon, OX11 ORA.

Dr. R. Fletcher, Dept. of Mathematics, University of Dundee, Dundee, DD1 4HN.

Lord Flowers, Rector, Imperial College, South Kensington, London, SW7.

Dr. D.P. Hodgkinson, Theoretical Physics Division, Building 424.4, A.E.R.E., Harwell, Oxon, OX11 ORA.

Dr. J. Howlett, 20b Bradmore Road, Oxford.

Dr. J. Hubbard, Research Laboratory K34/281, I.B.M., 5600 Cottle Road, San Jose, California 95193, U.S.A.

Dr. A.M. Lane, Theoretical Physics Division, Building 424.4, A.E.R.E., Harwell, Oxon, OX11 ORA.

Dr. A.B. Lidiard, Theoretical Physics Division, Building 424.4, A.E.R.E., Harwell, Oxon, OX11 ORA.

Dr. W.M. Lomer, Directorate, Building 329, A.E.R.E., Harwell, Oxon, OX11 ORA.

Dr. W. Marshall, Deputy Chairman, Building 329, A.E.R.E., Harwell, Oxon, OX11 ORA.

Dr. M.J. Norgett, Theoretical Physics Division, Building 424.4, A.E.R.E., Harwell, Oxon, OX11 ORA.

Dr. R. Phillips, Science Research Council, Rutherford Laboratory, Chilton, Didcot, Oxon, OX11 OQX.

Prof. M. Powell, Dept. of Applied Mathematics & Theoretical Physics, University of Cambridge, Silver Street, Cambridge, CB3 9EW.

Dr. I.C. Pyle, 16 Clifton Dale, York, YO3 6LJ.

Dr. J. Rae, Theoretical Physics Division, Building 424.4, A.E.R.E., Harwell, Oxon, OX11 ORA.

Dr. K.V. Roberts, Experimental Division B, Culham Laboratory, Abingdon, Oxon, OXL14 3EA.

Mr. P. Schofield, Materials Physics Division, A.E.R.E., Harwell, Oxon, OX11 ORA.

Dr. A.M. Stoneham, Theoretical Physics Division, Building 424.4, A.E.R.E., Oxon, OX11 ORA.

Dr. J. Tait, Theoretical Physics Division, Building 424.4, A.E.R.E., Harwell, Oxon, OX11 ORA.

Prof.f W.B. Thompson, Dept. of Physics, B-019, University of California, La Jolla, California 92093, U.S.A.

FOREWORD

It is a great pleasure to write the foreword to the Special Progress Report for Theoretical Physics Division in commemoration of the Atomic Energy Authority's 25th Anniversary. An anniversary like this inevitably creates a great deal of nostalgia and I look back at the history of the Division and my own personal career with total astonishment because events which are still very real to me, in reality occurred so long ago.

I was recruited to Harwell by Brian Flowers, who was then the Division Head, just 25 years ago and I can remember being given two mathematical tasks on my very first day. The first was to check Bill Thompson's algebra on some complicated calculation of plasma instability: I never did succeed in that because Bill had, and probably still has, a unique ability to use c.g.s. and m.k.s. units randomly and still arrive at (approximately) the right answer. The second problem was to decide whether I wished to retire on the old pension system or the new one. I was such a young man then, it had never occurred to me until then that anyone ever thought about pensions.

The following ten years were my happy ones. I even enjoyed being Division Head and I took a pride in gathering into the Division as many bright people as I could find. I left the Division on a sad day in 1966 and I have watched its progress and evolution with great interest ever since. This list of special reviews is fascinating to me and I hope it will be of interest for both nostalgic and scientific reasons to many other theoreticians.

I regularly threaten to return to scientific research and the Theoretical Physics Division but the present Division Head tells me that my publication record in recent years is not aceptable to him. I can understand that because reading these special reviews brings home to me the impact the Division has had on science and the nuclear programme. I wish it the best of luck for the future.

Walter Marshall.

Chapter I

INTRODUCTION

A. B. Lidiard

> The natural view to take of the world is that there are things which change; for example there is an arrow which is now here, now there. By bisection of this view, philosophers have developed two paradoxes. The Eleatics said that there were things but no changes; Heraclitus and Bergson said there were changes but no things....
> Bertrand Russell, History of Western Philosophy.

As one of the founder Divisions at Harwell, Theoretical Physics Division has seen many changes and has itself changed often and quite substantially. Only three present members of the Division were here at the beginning of our 25 year period; they are John Tait (1947-), Tony Lane (1953-) and myself (1954-7 and 1961-). The attached Table provides a brief chronology of the Division. This shows that the technical nature of its work has altered greatly over the years. The articles which follow review the changes and developments in many of the fields which the Division has contributed to. Each is a brief scientific review made from a personal or local point of view (which is why we call this a special progress report). The Division has not, of course, been involved with all these fields throughout the 25 years of the Atomic Energy Authority's existence. My purpose in this Introduction is to describe briefly the background to these changing programmes. We can usefully consider the three decades separately, although we refer first of all to the early, pre-Authority years because what happened then sets the scene in 1954.

Pre-Authority Years[1]. Harwell was the first atomic energy laboratory in the U.K., set up in 1947 with Dr. (later Sir) John Cockcroft as

Director. The first work on reactor design, weapons design, chemical processing and fuel fabrication was done here and the earliest low power graphite piles (GLEEP and BEPO) were built here. These practical tasks nevertheless required a lot of physics and mathematical support and this largely determined the initial role of Theoretical Physics Division. But other establishments at Springfields, Risley, Windscale, Capenhurst and Aldermaston were soon set up to bring these developments to fruition on an industrial scale. Therefore, as Harwell entered the 1950's much of the demand for immediate calculational support was moving away to these other centres. Nevertheless, it was recognized that the scientific understanding of many of the underlying physical processes was poor and it was therefore considered essential to have groups of theoreticians to develop models of nuclear reactions, to evaluate the experimental results and to design new accelerators. Computational techniques had also to be developed to serve these theoretical and other programmes. Shortly afterwards, two other topics were introduced into the Division, namely plasma physics, as part of the project to produce controlled thermonuclear reactions (fusion reactions as they are now called), and solid state physics, with the aim of giving theoretical support to the metallurgists.

The 1950's. Thus when the Authority was formed in 1954 the Division contained six groups: (i) Nuclear Theory (Tony Skyrme), (ii) Neutron Transport Theory (John Tait), (iii) Plasma Physics (Bill Thompson), (iv) Accelerator Design (Bill Walkinshaw), (v) Solid State Physics (Mick Lomer) and (vi) Computing (Jack Howlett). The 1950's as a whole were a period of optimism and growth in scientific laboratories almost everywhere - for the wartime successes of engineering guided by scientific research (e.g. radar and tha atomic bomb) had convinced people that continued technical success depended upon a strong scientific base. One consequence of this growth was that by the beginning of the 1960's Theoretical Physics Division had donated more than half of the six groups which it contained in 1954 to new laboratories. In 1956 the Atomic Energy Establishment was founded at Winfrith and the Neutron Transport Theory Group of T.P. Division subsequently evaporated. In 1957 the Rutherford High Energy Laboratory was established next door at Chilton and the Accelerator Group under Bill Walkinshaw joined it. The Culham Laboratory was set up in 1960 and the Plasma Theorv Group under Bill Thompson moved there in 1961. In the same year the Atlas Computing Laboratory was established alongside the Rutherford Laboratory with Jack Howlett as Director. Many of his Computing Group moved over with him.

The 1960's. This outward movement of the more directly applied work called for new activities and new Groups to take the place of those which had gone. In 1961 I returned to Harwell from academic life to set up the Crystal Defects and Radiation Damage Group and in 1962 the Division established an Atomic Theory Group under John Tait and a Neutron Group under Peter Schofield. We can point to two main influences in the 1960's, an organisational one, i.e. the need for a new role for Harwell, and a technical one, namely computers. In seeking a new role Harwell at first moved in an academic direction, both in its research programmes and in its proposals for new facilities, e.g. the High Magnetic Field Laboratory and the High Flux Beam Reactor (neither of which came off)*. T.P. Division's basic work

*Although it should be noted that the Variable Energy Cyclotron, which had a more applied function, namely the simulation of the behaviour of materials in reactors, did.

expanded rather quickly and its reputation in nuclear theory, neutron scattering, atomic cross-sections, magnetism, many-body theory, dislocations and defects climbed rapidly. The expansion in basic science at Harwell was halted by changes in national policy (and fortunes) in the mid-60's. Instead, Harwell now moved towards industry and sought to develop non-nuclear applications of nuclear techniques and know-how and to 'diversify' its programme; which is the course it is still following. By this time, however, it had acquired its own major computer, an IBM 360-65, and it was the Division's task to operate, maintain and develop the Harwell computing system. This acquisition meant a steady growth of Alan Curtis's Mathematics Group - as it was initially and Mathematics Branch as it became - and also that our own diversification was initially into computing applications (optimization methods and operations research, real-time systems, etc.). Individuals moved into these applications from nuclear physics, atomic physics and solid state physics. By 1970 these activities had grown so substantially and become so diverse that it was necessary to subdivide the Mathematics Branch into no less than four groups, (1) Numerical Analysis (Mike Powell), (2) Operations Research (Ian Cheshire), (3) Real-Time Computing (Ian Pyle) and (4) the Central Computer (Don Sadler).

The proximity of the computer itself, of computer experts and of numerical analysts had its influence on the more traditional physics activities. In particular the modelling of solids and liquids was becoming established quite firmly by the end of the decade. But the initial effects of the diversification programme were somewhat depressing to the morale of the physicists; nuclear, atomic and solid state theory were all obliged to contract, although the effects were made less severe by the continued presence of visitors and other attachments who still came as a consequence of the reputation established in the early 1960's. This side of the Division therefore turned towards the major experimental projects - reactor materials, reactor safety, heat transfer and fluid flow - and became increasingly involved in a way which was more akin to the style in the 1950's than to that of the early 1960's. The aim was to ensure that these projects obtained the best possible assistance from theory by way of understanding, interpretation of experiments and the construction of theoretical models. This meant that we often had to wrestle with poorly characterized systems and to make our own decisions on the important features of these systems; this is where the theoretician's understanding of fundamental principles and his (her) particular 'Weltanschauung' may differ from the experimentalist's and make an especial contribution to the success of a project. The end of the decade saw a growing self-confidence in the Division in its ability to tackle these complex and applied problems successfully, a confidence which steadily grew during the next ten years.

The 1970's. The Division thus entered the '70's broadly balanced between applied mathematics and computing activities, on the one hand, and theoretical physics activities on the other. Growth in these computing activities had not, of course, occurred in Theoretical Physics Division in isolation - as the articles by Roger Fletcher, Mike Powell and Ian Pyle clearly show. One consequence was that in 1973 the major computer activities at Harwell were brought together to form Computer Sciences and Systems Division and that Mathematics Branch became part of it. Further departures were to come. In 1975 Phil Hutchinson's section of the Physics of Fluids Group (under Peter Schofield) left to help found the new Thermodynamics

Division, while Peter Schofield himself transferred to Materials Physics Division soon afterwards to take responsibility for all neutron beam work. In the same year John Hubbard left us for the sunshine of California. The rest of the 70's saw the Division taking up new tasks (e.g. problems in radioactive waste disposal, the theory of ultrasonic methods of non-destructive examination and the modelling of oil wells) and expanding its effort on some existing ones. It enters the 80's with a very wide programme indeed.

After these many changes of personnel and programme one could reasonably ask 'What is Theoretical Physics Division?'. Of the many replies one could give I think it is correct to emphasize the talent, professionalism and loyalty of the inidividuals within it, the coherence and confidence of the whole and its style of working, fundamental and powerful but with its feet on the ground. These characteristics are well illustrated in the detailed articles which follow.

I am very grateful to those past and present members of the Division who have taken time to write this valuable collection of reviews of fields which either are or have been its province in the past 25 years. I would also like to acknowledge the essential part played by my secretary, Mrs. Marjorie Owen, in producing this volume.

(1) Margaret Gowing (assisted by Lorna Arnold), Independence and Deterrence, Britain and Atomic Energy 1945-1952 (Macmillan, London, 1974) 2 vols.

Table

1954 - Atomic Energy Authority created.
- Sir John Cockcroft is Director of Harwell.
- Head of Theoretical Physics Division
Dr. B.H. Flowers (now Lord Flowers)
Groups: Nuclear Theory (T.H.R. Skyrme)
: Neutron Transport (J.H. Tait)
: Plasma Physics (W.B. Thompson)
: Accelerator Design (W. Walkinshaw)
: Solid State Physics (W.M. Lomer)
: Computing (J. Howlett)

1955 - T.P. Division moves from Building 329 to the new 8.9.

1956 - Neutron Transport Theory Group moves to A.E.E., Winfrith.

1957 - Accelerator Group moves to Rutherford High Energy Laboratory.

1958 - Dr. B. Schonland becomes Director of Harwell.

1958 - Dr. W.M. Lomer becomes Head of T.P. Division.

1960 - Dr. W. Marshall becomes Head of T.P. Division.

1961 - Dr. F.A. Vick becomes Director of Harwell.
- Plasma Physics Group moves to the Culham Laboratory.
- Dr. J. Howlett becomes Director of the Atlas Laboratory.
- Applied Mathematics Group re-formed (A.R. Curtis).
- Crystal Defects and Radiation Damage Group formed (A.B. Lidiard).

1962 - Atomic Theory Group formed (J.H. Tait).
- Neutron Physics Group formed (P. Schofield).

1963 - Dr. R. Spence becomes Director of Harwell.

1965 - Science and Technology Act empowers the Authority to do non-nuclear work ('Section 4').

1966 - Dr. W. Marshall becomes Deputy Director of Harwell.
- Dr. A.B. Lidiard becomes Head of T.P. Division.
- Applied Mathematics Group redesignated Mathematics Branch.

1967 - The IBM 360/65 is commissioned.

1968 - Dr. W. Marshall becomes Director of Harwell.
- Dr. R.J.N. Phillips' High Energy Physics Section of the Nuclear Theory Group joins R.H.E.L.

1970 - Four new groups formed in the Mathematics Branch:
(i) Numerical Analysis (M.J.D. Powell)
(ii) Operations Research (I.M. Cheshire)
(iii) Computer Systems (I.C. Pyle)
(iv) Central Computer (D. Sadler)

1973 - Mathematics Branch joined the new Computer Sciences and Systems Division.

1975 - Dr. P. Hutchinson's Heat Transfer Section of the Neutron and Liquid Physics Group joins the new Thermodynamics Division.

1976 - Dr. L.E.J. Roberts becomes Director of Harwell.

1979 - T.P. Division moves from Building 8.9 to the new 424.4.

Chapter II

TWENTY FIVE YEARS OF FUNDAMENTAL THEORY

J. S. Bell

This period has brought no revolution in fundamental physical theory. In the absence of gravitation, Lorentz invariance remains a requirement on fundamental laws. Einstein's theory of gravitation inspires increasing conviction on the astronomical scale. Quantum theory remains the framework for all serious effort in microphysics, and quantum electrodynamics remains the model of a fully articulated microphysical theory, completely successful in its domain. However, a number of ideas have appeared, of great theoretical interest and some phenomenological success, which may well contribute to the next decisive step.

As regards Lorentz invariance, early in this period(1) there emerged a new consequence which came to be known as the 'PCT theorem'. In conventional local field theories Lorentz invariance, by an analytic continuation(2), automatically implies a certain discrete symmetry involving inversion in space (P), reversal of charges (C) and reversal in time (T). This result came into prominence a little later when P was found experimentally to be disrespected at the weak interaction level, and into further prominence later still, in the mid-sixties, when CP and T symmetries were found to be disrespected at a somewhat lower level. The violation of parity was rather quickly and very fruitfully incorporated into weak interaction phenomenology. The violation of CP and T remains even now something of an undigested marginal curiosity, but is easily accommodated in conventional theories(3). Any experimental violation of PCT would be much more embarrassing. None has appeared yet.

As regards gravitation, Einstein's theory has passed a series of increasingly sophisticated experimental tests[4]. But its prediction of gravitational radiation remains unverified. Such radiation seemed to have been detected in one experiment, but did not show up subsequently in others[5]. The observed damping of a certain binary star seemed to require the existence of such radiation, but at the time of writing it is not clear that other damping mechanisms have been sufficiently allowed for[6].

The theory of gravitational collapse, and black hole formation, has been much developed. Theorems on the inevitability of singularities raise hopes for a natural and neat beginning and end to things[7]. But these are theorems in classical theory, and there are indications that the singularities will not survive quantization[8]. A definite decision requires first an agreed synthesis of general relativity and quantum theory - not yet available despite much effort[9]. One remarkable quantum effect has, however, been established by convincing semi-classical reasoning - the energy of a black hole leaks away in the form of black body radiation[10]. Black holes are less permanent than had been thought.

In microphysics, quantum electrodynamics appears now, even more than at the beginning of this period, a miracle of precision[11]. In the light of new experiments it seems valid, for electrons and muons, right down to the level at which strong interactions inevitably intrude - by way of virtual hadrons in vacuum polarization. Early in this period strong interactions were themselves envisaged as governed by analogous quantum field theories. But the experimental multiplication of 'elementary' hadrons, and despair of doing reliable calculations with large coupling constants, led to a movement of many theorists away from field theory. Setting aside electromagnetism (as 'part of the equipment of the observer') they theorized separately about strong interactions and massive particles in terms of S-matrix concepts only, invoking analyticity assumptions and 'bootstrap' hypotheses - i.e. that all hadrons are on the same level and each is somehow made up of the others. This phase seems over, and quantum field theory is again the centre of attention. One pointer back to field theory was the phenomenological success of a simple composite model of hadrons, in terms of hitherto not directly observed, and perhaps not directly observable, elementary 'quarks'. Another was the success of certain sum rules and low energy theorems (collectively 'current algebra') of field theoretic inspiration. Finally, and most decisively, was the successful enlargment of quantum electrodynamics to cover also weak interactions(12) - in a way holding out hope for the further incorporation of strong interactions on similar lines(13).

In this last development the ideas of 'gauge' symmetry and of 'hidden' symmetry were vital(14). Quantum electrodynamics permits certain symmetry operations to be performed independently at different space-time points, the resultant mismatch being taken up by a corresponding transformation of electromagnetic potentials. This is the meaning of 'gauge' symmetry; quantum electrodynamics exemplifies the simplest 'Abelian' case. Gauging non-Abelian symmetry groups involves replacing the photon by more than one massless 'gauge boson' - and the extra bosons mediate interactions other than the electromagnetic. However, there is not, in fact, any great symmetry apparent between electromagnetic and other interactions in nature. Therefore the hiding (or, less appropriately, 'spontaneous breaking') of the symmetry is vital indeed. In fact the ground state of a system (the vacuum here) need not exhibit all the symmetry of the fundamental equations. The Heisenberg ferromagnet, with magnetisation pointing in some one of infinitely many equally suitable directions, is the traditional example. It has been found possible to contrive such a disymmetry of the vacuum so that the extra bosons become effectively massive, thus distinguishing in range and effective strength between the various interactions.

While these ideas were attractive in themselves, and permitted an elegant marriage of electromagnetic and weak interaction theory, what really attracted attention was the demonstration(15) that such theories are 'renormalizable'. That is to say that the infinities which plague all local quantum field theories are under control in the same sense as in quantum electrodynamics. It has to be noted, however, that the phenomenological success, in weak interactions, so far involves only the lowest order and low energies. The characteristic features of renormalizable theories will appear only in higher orders or at higher energies. Even the (rather massive) bosons supposedly mediating the weak interactions remain to be seen.

As a result of these developments an immense and continuing effort has been devoted to non-Abelian gauge theories. A number of striking results have been obtained, some of which may or may not be relevant for the 'confinement' of quarks in composite systems, i.e. for their non-appearance as free particles. Among the most beautiful and surprising of these results was the discovery in some such theories of magnetic monopoles(16) - as perfectly regular solutions of the classical differential equations.

Magnetic monopoles are the most striking examples (in elementary particle theory) of 'solitons' - permanently compact solutions of the classical equations. Even before quantization they represent 'particles' of some kind. Such mathematical objects and their possible relevance to elementary particle physics were considered only by a couple of pioneers at the beginning of this period(17). Only much later have they been more widely studied. It has been conjectured that the 'bag' confining quarks may be a related object. And the early conjecture that certain solitons, in certain model boson field theories, are effectively fermions, has since been brilliantly demonstrated(18). The old confidence no longer holds that the fundamental fields must include a fermion field.

Among symmetries, available for gauging or not gauging, a quite new species(19) has appeared in this period - so-called 'supersymmetry'. Bose and Fermi fields, or particles of integer and half-integer spin, appear here together in a given symmetry multiplet. Since these fields have different Lorentz transformations, and respect commutation and anticommutation relations respectively, this is no trivial addition to the previous list of 'all possible symmetries'. Its appearance should perhaps be used as a cautionary tale, illustrating the limited human ability to list all possibilities - as also indeed could the final appearance of a renormalizable theory of weak interactions. Supersymmetry has not yet been found relevant in phenomenology, but is of very great theoretical interest. The supersymmetry transformations, mixing with different Lorentz transformations, necessarily involve space and time in a non-trivial way. Related to this, the gauging of supersymmetry can generate, among others, a gauge boson of spin 2 that can be used as the 'graviton'.

In this way certain 'extended supergravity' theories have been constructed[19] which exhibit - if suitable spontaneous breakdown is anticipated - something like a unification of gravitational with weak, electromagnetic, and strong interactions. At this time the most elaborate version is not quite rich enough for phenomenology. But it has remarkable properties as regards divergences. In general, field theories involving spins other than 0, $\frac{1}{2}$ and 1 have much worse infinities and the renormalization philosophy does not work. In 'extended supergravity' theories there are miraculous cancellations at the one and two loop levels, such that renormalization does work. What happens in higher orders is not known.

It may be that from such considerations will emerge a theory of gravitation which is as satisfactory as could be with perturbation theory about flat space-time. However, for cosmological applications something more than that will be required. It may also be that quantum cosmology will require a solution of the infamous 'interpretation problem' of quantum mechanics. Quantum mechanics is still taught as giving only probabilities for 'results of measurements' on the given 'system'. When the 'system' is the universe, where is 'measuring' equipment to be found? And where is the 'measurer'? There has been little progress with this 'interpretation problem' - whose very existence is denied by many. Among those who see it, some (including Einstein) have considered adding extra variables (to the quantum mechanical description of the system) for the probabilities (given by the wave-function) to be about. Einstein hoped that such variables might restore not only objectivity, but also local causality, and perhaps determinism. At the beginning of this period Von Neumann's famous 'impossibility proof', as regards the restoration of determinism, was still generally considered to be important. One small progress is that the axiomatic basis of this theorem is now generally seen as arbitrarily and unreasonably narrow. This would have been seen sooner if the provers of theorems had paid more attention to the builders of models. At the same time, however, the non-locality of quantum mechanics has been more explicitly and quantitatively demonstrated. Any extra (or so-called 'hidden') variables can only bring this non-locality into greater prominence. It seems therefore that Einstein's hope, of embedding quantum theory in a locally causal theory defined by partial differential equations in ordinary space-time, is no longer tenable. The situations which are critical in this matter do not involve extremes of energy, distance or time. But they are experimentally delicate, and even the new experiments of the last decade are far from the critical ideal. As far as they go they support quantum mechanics rather than locality(20).

1. G. Lüders, Dan. Mat. Fys. Medd. 28, No. 5 (1954).

2. J.S. Bell, Proc. Roy. Soc. A231, 479 (1955).

3. J. Ellis, M.K. Gaillard and D. Nanopoulos, Nucl. Phus. B109, 213 (1976).

4. C.M. Wills, in General Relativity, ed. S.W. Hawking and W. Israel (Cambridge U.P. 1979) p.24.

5. D.H. Douglas and V.B. Braginsky, in General Relativity (see Ref. 4) p.90.

6. J.H. Taylor, L.A. Fowler and P.M. McCulloch, Nature, 277, 437 (1979).

7. C.W. Misner, K.S. Thorne and J.A. Wheeler, in Gravitation (Freeman, San Francisco 1979) p.1196.

8. J. Demaret, Nature 277, 199 (1979).

9. B.S. De Witt and S.W. Hawking, in General Relativity (see Ref. 4) pp.680, 746.

10. S.W. Hawking, Commun. Math. Phys. 43, 199 (1975) and P.C.W. Davies, Rep. Prog. Phys. 41, 1313 (1978).

11. J. Bailey et al, Nucl. Phys. B150, 1 (1979).

12. M.K. Gaillard, Nature, 279, 585 (1979).

13. M.K. Gaillard, Comments in Nucl. and Part. Physics, to appear.

14. S. Coleman, in Laws of Hadronic Matter, ed. A. Zichichi (Academic Press, New York 1975) p.173.

15. G. 't Hooft, Nucl. Phys. B79, 276 (1974).

16. P. Goddard and D. Olive, Rep. Prog. Phys. 41, 1357 (1979).

17. T.H.R. Skyrme, Proc. Roy. Soc. A247, 260 (1958); A262, 237 (1961); Nucl. Phys. 31, 556 (1962) and J.K. Perring and T.H.R. Skyrme, Nucl. Phys. 31, 550 (1962).

18. S. Coleman, in New Phenomena in Subnuclear Physics, ed. A. Zichichi (Plenum, New York 1977) p.297.

19. B. Zumino, Lecture Notes in Physics 100 (Springer-Verlag, Berlin 1979).

20. J.F. Clauser and A. Shimony, Rep. Prog. Phys. 41, 1881 (1978).

Chapter III

25 YEARS OF HIGH ENERGY PHYSICS

R.J.N. Phillips

High energy physics is about sub-nuclear particles, and the first thing to notice is how their numbers have multiplied. In 1954 a few dozen were known, including strange meson and baryon states, but during the sixties partial-wave analyses of scattering data uncovered a huge spectrum of excited states. These had the additive quantum numbers that were already known (charge, baryon number, strangeness) and decayed rapidly to a few long-lived or stable components. Since 1974, however, new quantum numbers have appeared too (charm, beauty, ...) in new, heavy and relatively long-lived particles. There is also a new lepton tau with its own neutrino, similar to the electron and muon but apparently quite independent. The population has thus exploded both in the number of sub-nuclear species (labelled by quantum numbers) and in the number of energy levels within each species(1).

This abundance of particles has been remarkably well explained by the quark model(2), which was evolved in the sixties. A few basic spin-$\frac{1}{2}$ entities called quarks are postulated and the known strongly interacting particles (hadrons) are interpreted as quark-antiquark or as three-quark states. Strong interactions conserve all quarks, so with n types (flavours) of quark we get n additive quantum numbers. If interquark forces do not depend on flavour we get an SU(n) symmetry for the strong interactions. The spectroscopy of the fifties and sixties can all be explained by three flavours of quark (labelled u,d,s with electric charges 2/3, -1/3, -1/3) that provide three additive quantum numbers plus an approximate SU(3) symmetry -

broken by ascribing unequal masses to the quarks. There is, however, a small but significant complication; the observed spectrum calls for a three-quark ground state that is totally symmetric in spin and space co-ordinates, apparently violating the spin-statistics theorem. The solution is to postulate a new internal degree of freedom (colour). Each quark belongs to a triplet representation of an SU(3) colour symmetry, whereas the observed hadrons are supposed to be colour singlets; the resulting three-quark ground states are antisymmetric in their colour indices and overall antisymmetry is restored.

In sub-nuclear spectroscopy quarks never appear in isolation, and indeed they might seem to be purely a mathematical convenience. The first indications of a possible deeper significance came with the identification of orbitally excited states: quarks began to look much more like real particles. Since 1968 a series of scattering experiments involving big momentum transfer have pointed even more strongly to a particle-like identification. The proton is a rather soft and extended object in sub-nuclear terms: it cannot absorb big momentum transfer - it disintegrates - and the elastic ep ep scattering cross section therefore falls dramatically at large Q^2 (invariant momentum-transfer squared). Measurements show, however, that inelastic scattering ep eX does not fall in the same way; somewhere among the disintegrated proton's constituents there seem to be hard, pointlike objects that can absorb big Q^2. May they, in fact, be quarks? Detailed analyses indicate that these objects have spin of $\frac{1}{2}$, and their electromagnetic and weak interactions correspond closely to what we expect for quarks. They seem to collide like particles, recoil like particles, but in the final stages nevertheless conspire to group themselves into qq and qqq systems so that bare quarks are never seen. Apparently there is a selection rule allowing only colour-singlet isolated systems. However, the energy and momentum of a fast recoiling quark are transmuted into a jet of hadrons that can be identified. Quarks give a meaning and a dynamics to jets.

Weak interaction theory too has changed radically. The discovery of parity and charge-conjugation violation for weak interactions in 1956 led quickly not to confusion but rather to the universal V-A current-current interaction. This form suggests a gauge theory analogous to quantum electrodynamics (Q.E.D.), where the weak interactions are transmitted by charged and neutron bosons analogous to photons, and these neutral bosons predict new "neutral current" transitions. Neutral currents were duly observed in 1973, but a theory based on u,d and s quarks alone also predicted s d transitions that were not observed. The most economical solution was a more symmetrical theory containing one more charge 2/3 quark, c. When this quark too was discovered in 1974 it seemed a clear vindication of gauge-theoretical principles, and most of the present thinking is now in this direction. The present standard SU(2) x U(1) gauge model with spontaneous symmetry breaking unites weak and electromagnetic interactions in a single formalism: it successfully predicts ten independent neutral current matrix elements in terms of a single parameter(3).

Since gauging works for these electro-weak processes, what about strong interactions? Colour SU(3) could be a gauge symmetry: if so, there are coloured gauge bosons (gluons) analogous to photons, transmitting forces that could be the basic strong interactions. This quantum chromodynamics (Q.C.D.) is being studied intensively and seems the most promising present approach to strong interactions. There are important differences from quantum electrodynamics (Q.E.D.) since gluons not only couple to colour sources but are themselves coloured, and couple to each other. The theory is asymptotically free; the renormalized coupling constant depends on the relevant momentum scale, and for very high momenta it tends logarithmically to zero. This offers a new area of application for perturbative field theory, successful in Q.E.D. but previously ineffectual for strong interactions. Conversely, at low momenta (or large distances) the theory has unfamiliar divergences that may conceivably explain colour confinement - the apparent requirement that isolated systems be colour singlets(4).

There is even hope of combining strong and electro-weak forces in a single Grand Unified Gauge Theory with spontaneous symmetry breaking. Such a theory would explain the quantization of charge. It would also treat leptons and quarks on the same footing, putting them in the same super-multiplets; hence some of the new gauge bosons would necessarily transmute quarks into leptons and allow nuclear matter itself to decay. For example, the proton might decay via $p \rightarrow \gamma\gamma e^{+}$. Candidate theories of this kind put the proton lifetime near 10^{30} years, close to the present experimental limits on this quantity(5).

We talk mostly of quarks and gluons nowadays, but they are not the only possible language for hadron physics. In the framework of S-matrix theory there is a symbiosis of particles and forces. Forces can generate particles, as bound or resonant states of other particles; on the other hand, the exchange of particles generates forces. So particles generate forces which generate particles, and in principle there might be a "bootstrap" solution in which all particles generate each other. The sixties saw determined attempts to construct dynamical bootstrap solutions, with increasingly sophisticated treatments of particle exchange (fixed poles, Regge poles) and methods of relating exchange forces to the particles they create (N/D approximation, duality). This approach implies sub-nuclear democracy - all particles are equally basic - and is the antithesis of the quark model. However, no complete workable dynamics was found, although there was quite extensive success in correlating many previously disconnected areas[6].

These paragraphs have sketched just a few of the new ideas, arbitrarily omitting many others, but they illustrate astonishing changes. Twenty-five years ago who would have guessed at parity violation, or an invisible quark-gluon infrastructure, or proton instability? What will the next quarter-century astonish us with?

1. Review of Particle Properties, Phys. Lett. 75B (1978).

2. F.E. Close, "An Introduction to Quarks and Partons", Academic Press, 1979.

3. See talks by C. Baltay and S. Weinberg in Proceedings of the 19th International Conference on High Energy Physics, Tokyo, 1978 (Physical Society of Japan, 1979).

4. See, e.g., M.K. Gaillard talk at European Physical Society International Conference on High Energy Physics, Geneva, 1979.

5. See, e.g., J. Ellis talk, ref. 4.

6. G. Veneziano, Phys. Rep. 9, 199 (1974).

Chapter IV

NUCLEAR REACTIONS

A. M. Lane

1. Relevance of Nuclear Reaction Studies to Atomic Energy

Nuclear energy is released by reactions that occur between colliding nuclei. Part of the internal energy of the colliding nuclei is converted by the reaction into kinetic energy of the final nuclei, and therefore into usable heat energy. This is true both for fission reactions (typically resulting from neutron-uranium collisions) and fusion reactions (arising from collisions of certain light nuclei). Each reaction at each colliding energy has a characteristic cross-section which describes the probability that the reaction process occurs. Of all the physical parameters that enter into the design of a reactor, the most directly relevant and crucial are the reaction cross-sections that determine the energy-releasing processes. If cross-sections are small, no net energy release occurs; if they are large, reactor controls must be designed to restrict the energy release to manageable amounts.

This essential role of nuclear cross-sections in reactor design accounts for the extensive theoretical and experimental studies in these quantities at nuclear energy research stations. When possible, cross-sections are measured by experiment using focussed beams of charged particles from accelerators or collimated beams of neutrons from various kinds of sources. Sometimes experimental measurement is not feasible; suitable target materials may not be available, or beam resolution may be inadequate. In such cases, theoretical studies are needed to give a guide to the magnitude of the required cross-sections. Even if such direct support were not needed, theory still has a role in a healthy research programme by providing physical interpretation of the results of the experimentalists.

At Harwell over the last twenty-five years, both experiment and theory have struck a balance between work on cross-sections of direct practical relevance to reactors and on those relevant to an overall understanding of reaction processes.

2. Pre-1954 Background

There are two kinds of reaction process that dominate nuclear reactions, the so-called "compound nucleus" and "direct" processes.

2.1 Compound nucleus process

In this case, the two colliding nuclei fuse completely to form an intermediate state which subsequently decays. This two-step process was first proposed by Bohr in 1936, and was believed for many years to be the dominant mechanism for nuclear reactions. This belief was founded partly on the experimental observation in neutron capture of very intense narrow resonances in the excitation curves. . Such resonances are direct evidence for the existence of long-lived intermediate states. Another source of the belief was the fact of very strong, short-range nuclear forces. This made it appear quite natural that any interaction between two colliding nuclei should lead to a complete merging of the two bodies. A key concept here is that of the mean free path of a nucleon against a collision with other nucleons. The nature of nuclear forces suggests a very short mean free path, with the implication that colliding nuclei fuse quickly into each other.

At low bombarding energies, the resonances (which correspond to unbound levels of the combined nuclear system) are widely spaced and narrow, so that at most one ie excited at a given energy. In 1938, a detailed theory was given of this kind of resonance excitation, and the resulting cross-section expression was called the Kapur-Peierls dispersion formula.

At higher bombarding energies, resonances become wider and more closely spaced, so they begin to overlap and eventually the cross-section becomes smooth. Nevertheless, it is composed of a dense set of resonances. In 1938, Bethe gave a theory for such situations in which he assumed that the amplitudes with which different states are excited are random. This assumption was in the spirit of the Bohr theory and resulted in an expression for the cross-section which was the product of two factors, one for each of the entrance and exit channels. This product form enshrined the independence of the formation and decay processes implied in the Bohr theory. (In fact, Bethe's theory was only semi-quantitative. It was not explicitly consistent with formal requirements like the unitarity of the scattering matrix. It is only in the period 1964-76 that rigorous derivations have been given of the product formula).

2.2 Direct processes

About 1950, in (d,p) reaction studies, features were observed that were in violation of expectations from compound nucleus theory. The angular distribution was found to be strongly peaked in the forward direction, in sharp conflict with the near-isotropy expected. (This isotropy is an expression of the compound nucleus concept that the decay process has 'no memory' of the formation process). The observed facts were quickly described by a new theory, that of "direct" processes (or "stripping" in the particular case of (d,p) reactions). In this description there is a single interaction between the colliding nuclei that directly causes a transition to the final nuclei, without the intervention of an intermediate state.

Since the establishment of this "direct" process in (d,p) studies, counterparts have been found in almost all kinds of reaction. Depending on the reaction and on the energy, one process or the other may dominate, but in general both are present.

In the simplest descriptions, one pretends that the processes are independent and that the cross-sections are calculated separately, then added. In fact, provided that cross-sections are averaged over resonances, it is formally correct to add them. However, this does not mean that they can be calculated independently; indeed they affect each other in a complicated way which has only recently received adequate theoretical treatment.

This brief survey gives the background to the contributions to nuclear reaction theory that have occurred at Harwell over the last twenty-five years.

3. Nuclear Reaction Theory at Harwell 1954-79

During this period, almost all developments in nuclear reactions were assisted by contributions from Harwell. These contributions include:

Optical Model for Nucleon Reactions (A.M.L.)

Nucleon-Nucleon Scattering (J.K.P.)

Alpha-Alpha Scattering (J.K.P.)

Theory of (d,p) Reactions with Coulomb-Effects (C.F.C.)

Neutron Capture Reactions (A.M.L., C.F.C., J.E.L.)

Calculation of Individual Resonance Widths (A.M.L.)

R-Matrix Theory of Resonance Reactions (A.M.L.)

Coulomb Scattering by Deformed Nuclei (C.F.C.)

Calculable Theory of Reactions (A.M.L.)

Reactions with Doorway States (A.M.L., J.E.L.)

Isospin-dependent Optical Potential for (p,n) Reactions (A.M.L., J.M.S.).

Reactions with Analogue States containing Fine Structure (A.M.L.)

Correlations between Partial Width Amplitudes of Resonances (A.M.L.)

Sum Rules for Spectroscopic Factors (C.F.C.)

Time-Reversal Violation in Reactions (C.F.C.)

Sum-Rules for Photonuclear Reactions (A.M.L.)

Threshold Anomalies (A.M.L.)

Theoretical Analysis of Proton Scattering at the Analogue of ^{208}Pb (D.W.)

Optical Model Analysis of Scattering Data (D.W.)

Hauser-Feshbach Analysis of Compound Nucleus Data (D.W.)

The initials are those of the relevant authors, i.e. C.F.C. = C.F. Clement, J.K.P. = J.K. Perring, A.M.L. = A.M. Lane, J.E.L. = J.E. Lynn, D.W. = D. Wilmore, J.M.S. = (the late) J.M. Soper.

4. Representative Selection of Harwell Work

Rather than trying to give an account of all the above items, a few items of central importance will be chosen and expanded upon.

4.1 The optical model

In describing reactions initiated in a nuclear collision process, the most basic features are the cross-sections for elastic scattering and for absorption. The absorption is composed of all reaction cross-sections, and, with the elastic cross-section, it forms the total cross-section. The Bohr model, with its very short mean free path, implies that the wave-function of relative motion of the colliding nuclei has only in-going components at the point where the nuclei meet, thereby enshrining the idea of strong absorption. However, the discovery of direct reactions in 1950 meant that the mean free path was not as short as that implied by this picture. Rather, one should relax the in-going-wave assumption to allow for the fact that sometimes the incident particle can return to the entrance channel. This means that the nucleus presents to the incident particle a potential which is not entirely absorptive, and therefore can refract as well as absorb, i.e. it is an optical potential.

The notion of a mean free path long enough to enable the particle to survive a transit across the nucleus caused some consternation in the face of the strength of nuclear forces. However, this feature was also implied by a concurrent development in nuclear structure, viz. the rise of the independent particle model, and therefore had to be accepted as fact. The key to the paradox is the Pauli Principle which effectively dilutes the collision power of nuclear forces by forbidding many transitions. For an incident nucleon, this effect is very strong at low energies where nearly all states allowed by energy and momentum conservation are forbidden. This means that the absorption part of the optical potential is strongly reduced. In 1955, a simple quantitative theory of this effect was given, and this fitted the data on absorption.

This discussion of the optical potential has been a semi-classical one, and has not mentioned the underlying quantum aspects embodied in resonances. In 1955, a theory was given which built the bridge between the phenomenological model, and the full microscopic description involving the fine-scale resonances.

4.2 Neutron capture theory

In a series of papers between 1957 and 1976, the theories of reaction mechanisms were applied to neutron capture. Because of the striking resonances at low energies, it had been assumed that neutron capture was exclusively a compound nucleus process. This is almost true at low energies, but less so at higher energies where the compound nucleus cross-section is sharply reduced by other competing decay channels. Even at low energies, between resonances, there is a small direct cross-section.

It was shown in 1957 that 14 MeV capture cross-sections could not be understood in compound nucleus terms, but that they required a direct mechanism. Later studies showed that the direct mechanism was not simply a matter of the incident particle making a radiative transition to a bound orbit, but was modified by collective effects that redistribute the radiative strength in energy. The result was the "semi-direct mechanism" in which the incident particle excites the giant dipole resonance while being scattered into a bound orbit, with the subsequent radiative decay of the giant resonance. This picture continues to be the standard one for discussion of fast nucleon capture.

In the low-energy region, direct capture rarely shows itself in the cross-section, being drowned by the compound nucleus cross-section. However, it shows itself dramatically in another form. As we saw in our introductory remarks, the two mechanisms cannot operate independently, but must affect each other. The most striking effect is that the presence of direct capture implies that there are correlations between the neutron and radiative widths of resonances. Over the years an impressive list of experimental cases of correlations has been accumulated, and these have been explained(5), at least semi-quantitatively, as a consequence of direct capture.

4.3 Reactions at doorway states with fine structure

If the compound nucleus process dominated reactions, then all aspects of reactions would be statistical, and the absorption cross-section in any channel would be a smooth function of energy. An early indication that other processes could occur was the discovery that broad resonances occurred in neutron absorption cross-sections, and these were fitted with the optical model. Since then, much more concentrated and dramatic peaks have been found in certain situations; prime examples are in fission channels (arising from so-called Class II doorway states) and in proton channels (arising from analogue states). In both cases, high resolution studies often show that the peak is composed of a large number of fine-structure resonances, whose parameters vary systematically with energy in order to give the peak observed with normal resolution. The interplay between the fine- and gross-structure has been a fascinating and profitable area of study(6), especially when there is a background with which the doorway state interferes (giving the so-called Robson Asymmetry).

In heavier nuclei, the fine-structure is not resolvable, and one has only the gross-structure, although this may often be observed in several channels. A good example is ^{207}Pb(p,p') where the analogue of ^{208}Pb appears in several inelastic cross-sections. The theoretical description of this situation has been given with an elaborate coupled-channels calculation(7), in which Coulomb effects give rise to the line-broadening of the analogue.

4.4 Sum-rules for spectroscopic factors

When (d,p) and (p.d) data on a given target are analysed with direct reaction theory, the result is a collection of a large number of spectroscopic factors for adding a neutron or a neutron-hole to the target. In certain cases, these may be fitted individually by appropriate theory, but often this is not possible because the theory (shell-model with configuration mixing) is prohibitively complicated. Then it is extremely useful to analyse the data with sum-rules(8). When the target has non-zero spin, there are many of these rules, with less parameters than rules. (This applies to energy-weighted, and non-energy-weighted rules). Thus there are, in effect, powerful consistency checks, as well as equations for the parameters. The values from this analysis are complementary to those obtained from fitting energy spectra, and greatly tighten up the whole process of fitting theories of nuclear structure to the available data.

1. A.M. Lane and C.F. Wandel, Phys. Rev. 98, 1524 (1955).

2. A.M. Lane, R.G. Thomas and E.P. Wigner, Phys. Rev. 98, 693 (1955).

3. A.M. Lane and J.E. Lynn, Nucl. Phys. 11, 646 (1959) and 17, 563 (1960).

4. C.F. Clement, A.M. Lane and J.R. Rook, Nucl. Phys. 66, 273 (1965).

5. A.M. Lane, Proc. of Int. Conf. on Neutron Capture Gamma-Ray Spectroscopy, Petten, Holland, Sept. 1974 (R.C.N. Petten, March 1975) p.31.

6. A.M. Lane, "Isospin in Nuclear Physics", (North Holland, Amsterdam, 1969) ed. D.H. Wilkinson, p.509.

7. S. Ramavataram, D. Wilmore and D.J. Edens, Proc. Int. Conf. on Properties of Nuclear States (Laval University Press, Montreal, 1969) p.761.

8. C.F. Clement, Nucl. Phys. A213, 469 (1973).

Chapter V

ELECTRONIC COLLISIONS

P. G. Burke

1. Introduction

The study of the collisions of electrons with atoms, ions and molecules has seen an enormous increase in activity, both theoretical and experimental, over the last twenty-five years. At the beginning of this period, the subject of electronic and ionic impact phenomena, which also includes thermal energy and high energy ion-atom, atom-atom and atom-molecule collisions, was comprehensively described within the covers of one book by Massey and Burhop(1). Twenty years later, in the early 1970's, Massey, Burhop and Gilbody, in the monumental second edition of this book, needed no less than five volumes to cover this subject(2). To-day, only five years later, it is impossible to contemplate ever bringing this subject together in this way again.

The growth in the subject has been caused by many factors. Perhaps the most important of these is the continuing and, indeed, ever increasing need for electron collision cross-sections in many applications. These include (i) the need for accurate rate coefficients to enable the electron densities and temperatures to be determined in astrophysical plasmas, such as stellar atmospheres, (ii) the role these processes play in gas lasers and (iii) the need for these cross-sections in understanding plasma fusion devices. On the experimental side, the development of new techniques and improved electronics has enabled a new generation of experiment to be carried out. No longer is it possible to measure only total or perhaps differential cross-sections. Instead, by using spin-polarized beams and various electron-electron and

and electron-photon coincidence techniques and laser pumping techniques, it is now often possible to measure the complete scattering amplitude, perhaps involving excited states. These experiments, of course, provide a much more stringent test of the theory and enable a deeper understanding of the regions of applicability of different theoretical models to be developed. Finally, on the theoretical side, new approaches have been developed and these, coupled with the vastly improved computing facilities which are now available, are enabling accurate cross-sections to be calculated in many cases of interest.

This review will concentrate on theoretical developments which have been made in the last twenty-five years. From this point of view it is often convenient to divide the energy range for the incident electron into low, intermediate and high energy regions. In the low-energy region the velocity of the incident electron is of the same order or less than the velocity of the electrons in the target which are taking an active part in the collision. In this region only a few target states can be energetically excited. The intermediate-energy region extends up to an energy where the velocity of the incident electron is typically about four times the velocity of the active target electrons. This is the hardest region to treat theoretically, since an infinite number of target states can be excited and also because ionizing collisions are possible. Finally, the high-energy region is characterized by the rapid convergence of perturbation theory and, at sufficiently high energies, the first Born approximation will usually, but not always, be applicable.

2. Low Energies

At low energies, the collision has many of the features of a bound state problem. The wave-function describing the collision can be accurately represented in terms of a sum of configurations, similar to the configuration interaction expansions used for bound-state calculations of atoms and molecules. This so called close-coupling expansion, introduced by Massey and Mohr[3] in the 1930's and developed by Seaton[4] and many others since, can be written for an electron incident on an N-electron target as:

$$\Psi(1,2 \ldots \ldots,N+1) = \mathcal{A}\sum_i \phi_i(1,2 \ldots,N)\, F_i(N+1) + \sum_i \chi_i(1,2,\ldots,N+1)\, a_i \qquad (1)$$

where the ϕ_i are a finite number of low-lying target states - those which are important in the collision - while the F_i describe the motion of the scattered electron and the χ_i are additional functions allowing for electron-electron correlation, which vanish unless all the electrons are close together. The total wave-function is antisymmetrized by the operator $\mathcal{A}$ in accordance with the Pauli exclusion principle. If the expansion (1) is substituted into the Kohn variational principle then coupled integro-differential equations are obtained for the radial parts of the functions F_i, which are coupled to linear equations for the coefficients a_i. Then, from the asymptotic form of the F_i the S-matrix and consequently the cross-sections for transitions between the states ϕ_i can be obtained.

An important development in the 1960's was the realization, particularly by Spruch and his colleagues[5], that, as for bound-state problems, the solution of the scattering problem, corresponding to expansion (1), satisfied certain bound principles. As an example, if the ground state of the target is known accurately, then the phase-shift calculated using expansion (1) is a lower bound on the exact phase-shift at low energies; the calculated phase will increase monotonically towards the exact phase-shift as the number of terms in the expansion is increased. Although the situation is more complex if the target states are imprecisely known, these bound properties of the solution are a major reason why reliable results can be obtained in this energy region.

Of course, in order to obtain these results, it is necessary to solve accurately and speedily the coupled equations resulting from expansion (1). In the case of electron scattering by light atoms and ions, where relativistic effects are unimportant, considerable progress has been made in this area in the last ten years and numerical methods and general computer program packages have been developed which now enable accurate cross sections to be calculated for an arbitary target(6,7). However, for electron scattering by heavy atoms and by molecules, the situation is less satisfactory. In the former case, relativistic effects, and, in the latter case, the multi-centred nature of the electron-molecule interaction, considerably complicate the form of the equations which must be solved. However, for both heavy atoms and molecules, recent theoretical developments in the use of L^2-integrable (square-integrable) wave-functions in collision calculations have opened up the possibility of modifying standard bound-state program packages to calculate collision cross-sections(8). These developments are being actively pursued by many groups and recent results, particularly in the case of electron-molecule collisions, are most encouraging(9).

It is also important to mention the fundamental role which resonances play in low-energy collisions. Although it was known in the 1930's that resonances could occur, it was not until the early 1960's that Fano(10) firmly focussed attention on their importance in photo-ionization processes. Shortly after this, the first detailed calculations and observations showed that resonances were a common feature of all low-energy collisions of electrons with atoms, ions and molecules(11). Since then their importance in such processes as di-electronic recombination, vibrational excitation and dissociative attachment has been clearly established. Nevertheless, certain basic questions still remain unanswered when the resonances involve interactions between more than two electrons, as is the case in the post-collision interaction discovered by Read and collaborators(12).

To conclude this discussion of low-energy collisions, the decisive influence which the development of multichannel effective-range or quantum-defect theories have had, particularly for electron-ion collisions, must be mentioned(13-15). These theories describe the behaviour of cross-sections in the neighbourhood of thresholds in the presence of a long-range, attractive, Coulomb interaction and they enable the resonance structure below threshold to be predicted from a knowledge of the scattering amplitude above threshold. As well as enabling a complicated resonance cross-section to be described in terms of a few parameters, the theory also provides a convenient way of interpreting experimental results.

3. Intermediate Energies

Turning now to intermediate energies, where an infinite number of target states can be energetically excited, it should first be noted that expansion (1) now has the wrong asymptotic behaviour. This is because only a finite number of target states can be included in the expansion and because there is no easy way of including the continuum states. Nevertheless, one approach which has had some success is based on this expansion in which some of the target states are replaced by suitably chosen 'pseudo-states', i.e. states which are not eigenstates of the target Hamilitonian. Instead these pseudo-states each represent an average in some sense over the complete set of target eigenstates. For example, Damburg and Karule(16) have shown how pseudostates can be constructed for atomic hydrogen, which represent the first-order distortion of the target in the field of the scattered electron, and this technique has now been extended to treat any atom or molecule. Using this approach, cross-sections for the excitation of atomic hydrogen have recently been calculated which are in good accord with experiment at intermediate energies. There is no basic reason why the same could not be done for more complex targets.

A number of other approaches involving some form of analytic continuation in the complex energy plane have recently been introduced at intermediate energies(8,17). The essential point about these approaches is that by a suitable choice of continuation, an L^2-integrable trial wave-function can be used, thus avoiding the difficulty of explicitly having to construct the asymptotic form. However, such approaches have so far been limited to elastic scattering from atomic hydrogen, and it remains to be seen how far it will be possible to extend them to describe inelastic electron collisions with complex atoms and molecules.

At this stage it is appropriate to mention the developments which have been made in the theory of ionization. It is almost exactly twenty-five years since Wannier(18) wrote his paper on the classical theory of ionization. In this theory he predicted that the ionization cross section near threshold had the form

$$\sigma \sim E^{1.127}$$

where E is the excess energy from threshold. For many years there was considerable controversy about this result with arguments made that a linear threshold law would be obtained using a fully quantal theory. The situation was clarified in the early 1970's by Peterkop(19) and Rau(20), who showed that Wannier's result was not inconsistent with quantum mechanics, and by Cvejanovic and Read(21), who studied the ionization of helium experimentally and obtained results which supported Wannier's law and which were definitely inconsistent with a linear behaviour at threshold. However, a completely ab-initio theory, which is capable of predicting the magnitude and shape of the ionization cross-section close to threshold, is still lacking.

4. High Energies

The theoretical understanding of electronic collisions at high energies has also seen many significant advances in recent years. Perhaps the most important of these is the realization that the first Born approximation does not always give the leading contribution to the cross-section. For example, inelastic collisions at large scattering angles are dominated by the second term in the Born series, in which the scattered electron interacts once with the nucleus, to give a large scattering angle, and once with a target electron to give excitation. Byron and Joachain(22) in a series of important papers have also made the point that, in order to obtain consistent results as the energy decreases from a very high value, it is necesssary to retain all terms in the Born series having the same energy dependence. In this way they find that the correction to the first Born approximation must involve terms from the real part of the third Born amplitude, as well as the second Born amplitude. They estimate the third Born contribution using an eikonal approximation and their results are in very satisfactory agreement with experiment.

5. Highly Excited States

The last topic which will be mentioned in this review is electronic collisions involving transitions between highly excited states. It is in this area where classical theories, which saw an enormous resurgence in interest following the work of Grysinski(23) some twenty years ago, have really come into their own. It is clear that quantal theories based upon expansion (1) are generally only appropriate up to a value of the principal quantum number n of about three or perhaps four, after which the number of states which need to be coupled becomes prohibitively large. However, highly excited states with an n value up to 105 have been observed in the laboratory, while radio-frequency observations of interstellar gas clouds have detected transitions between even higher values of n. For example, in 1965 an emission line at 5.4 GHz was observed coming from the Orion nebula and attributed to a transition between the levels n = 110 and n = 109 of atomic hydrogen. In a recent review, Percival and Richards(24) have examined regions of validity of various approaches which can be used to calculate transitions between these states. These range from purely classical calculations, using Monte Carlo methods, to methods based on Bohr's or Heisenberg's correspondence principles. The theory is now reasonably complete for energy-changing collisions in which the target is in a state within n greater than 5; however, this leaves a possible gap involving n about 4 or 5, where new approaches are still required.

6. Conclusion

In conclusion, although the field of electronic collisions has seen very substantial advances in the last twenty-five years, there are still many problems outstanding, and the field promises to continue to be as active and exciting in the future as it has been in the past.

1. H.S.W. Massey and E.H.S. Burhop, "Electronic and Ionic Impact Phenomena", 1st Edition (Oxford University Press, 1952).

2. H.S.W. Massey, E.H.S. Burhop and H.B. Gilbody, "Electronic and Ionic Impact Phenomena", 2nd Edition in five volumes (Oxford University Press, 1969-74).

3. H.S.W. Massey and C.B.O. Mohr, Proc. Roy. Soc. (London) A136, 289 (1932).

4. M.J. Seaton, Phil. Trans. Roy. Soc. (London) 245, 469 (1953).

5. Y. Hahn, T.F. O'Malley and L. Spruch, Phys. Rev. 134, B397 and B911 (1964).

6. K.A. Berrington, P.G. Burke, M. Le Dourneuf, W.D. Robb, K.T. Taylor and Vo Ky Lan, Comp. Phys. Commun. 14, 367 (1978).

7. M.A. Crees, M.J. Seaton and P.M.H. Wilson, Comp. Phys. Commun. 15, 23 (1978).

8. W.P. Reinhardt, Comp. Phys. Commun. 17, 1 (1979).

9. P.G. Burke, "Theory of Electron Molecule Collisions". (Lectures presented at the NATO Advanced Study Institute on Quantum Dynamics of Molecules, Cambridge University, Sept. 15-29, 1979, to be published by Plenum Press, 1980).

10. U. Fano, Phys. Rev. 124, 1866 (1961).

11. P.G. Burke, Adv. in Atom. Molec. Phys. 4, 173 (1968).

12. A.J. Smith, P.J. Hicks, F.H. Read, S. Cvejanovic, G.C.M. King, J. Comer and J.M. Sharp, J. Phys. B 7, L496 (1974).

13. M.J. Seaton, Proc. Phys. Soc. 88, 801 (1966).

14. M. Gailitis, Zh. Eksp. Teor. Fiz. 44, 1974 (1963)(Soviet Physics - JETP 17, 1328 (1964)).

15. U. Fano, Phys. Rev. A2, 353 (1970); ibid 15, 817 (1977).

16. R. Damburg and E. Karule, Proc. Phys. Soc. 90, 637 (1967).

17. F.A. McDonald and J. Nuttall, Phys. Rev. A4, 1821 (1971).

18. G.H. Wannier, Phys. Rev. 90, 817 (1953).

19. R. Peterkop, J. Phys. B 4, 513 (1971).

20. A.R.P. Rau, Phys. Rev. A4, 207 (1971).

21. S. Cvejanovic and F.H. Read, J. Phys. B7, 1841 (1974).

22. F.W. Byron, Jr. and C.J. Joachain, Phys. Rev. A8, 1267 (1973) and A9, 2559 (1974).

23. M. Gryzinski, Phys. Rev. 115, 374 (1959).

24. I.C. Percival and D. Richards, Adv. in Atom. Molec. Phys. 11, 1 (1975).

Chapter VI

INTERATOMIC COLLISIONS

J. S. Briggs

1. Relevance to Harwell Programmes

The study and understanding of interatomic collision processes has been closely interwoven with the development of nuclear physics and its technological applications since the early days of this century. However, it must be said that the relevance of such studies to nuclear power development has often been to provide insight into ancillary, but nonetheless vital, processes involved in the fission and fusion reactions. An example from the very earliest days of nuclear physics was the necessity to understand the stopping power of solids and liquids for fast charged particles (e.g. protons and α-particles) passing through them. Even to-day, these studies form a cornerstone of what is known as radiation physics, that is, the physical and biological effects of ionising radiations. However, stopping power is a gross measure of the interaction between heavy projectile ions and target atoms. Increasingly over the last fifteen years, which roughly marks Theoretical Physics Division's involvement in the problem, has emerged the need to unravel in much greater detail the inelastic events occurring in the close collision of two atomic systems. Specifically, the aim has been to develop the theory of the angular and energy distributions of the fragments (e.g. free electrons, charged ions and photons) emerging from the collision as a function of the collision velocity. The development of the detailed theory and particularly the involvement of Theoretical Physics Division in this development, has kept in step with the increasing relevance of atomic collision processes in the broadening of the Authority's interests over the last ten years or so. Particularly noteworthy are the applications in the field of ion-solid interactions, e.g., ion implantation, Auger and X-ray electron spectroscopy and ion-beam simulation of radiation damage processes. Potentially even more important is the increasing recognition of the important rôle of atomic collision processes in the fusion programme where α-particle heating, all manner of impurity energy-loss processes and neutral beam injection depend crucially upon the magnitude of cross-sections for electron capture and loss in heavy ion-atom collisions.

It is with this background that we come to discuss the steps in the development of the theory of inelastic atom-atom collisions and the impact of the particular contributions made by members of T.P. Division. It can be asked, and often is, just what problems remain in the field of atomic collisions where the forces are known to be purely Coulombic, where (apart from some aspects mentioned below) non-relativistic theory is sufficient and where the interaction with the radiation field can often be considered only in first-order perturbation theory. The answer to this lies in three aspects of the problem. One is the rather obvious fact that one is dealing with a quantum-mechanical few-body problem; that is to say, the number of particles (nuclei, electrons and photons) is, in general, neither so few to allow reduction to an effective one- or two-body problem, nor so great to manifest either statistical or 'many-body' behaviour, for example by the occurrence of collective excitations. The second is that the Coulomb force is notoriously difficult to handle both analytically and numerically. Being inversely proportional to the interaction distance, it is essentially of infinite range and so gives rise to the Coulomb-wave phase-shift which must be accounted for even in the 'zero-order' problem, i.e. outside the interaction region. For this reason, many of the proofs and manipulations of formal scattering theory become invalid for the Coulomb potential. Finally, because atomic cross-sections are often large (10^{-12} - 10^{-24} cm^2) by the standards of nuclear or particle physics and thus are easy to measure, the accuracy of the measurements imposes a similar accuracy upon the theory which seeks to explain them. For these reasons the theory of atomic collisions still presents unsolved problems, particularly in the case of ionisation and re-arrangement collisions.

2. Historical Development

The theory of atomic collisions concerns the calculation of energy and momentum transfer between nuclei and electrons interacting via the Coulomb force. Different energy and momentum transfers lead to different final bound or unbound states and the aim of the theory is to calculate cross-sections for such processes. Almost all atom-atom collisions lead to the emission of photons either during or after the collision and the frequency and angular dependence of this photon emission provides a sensitive test of collision theories.

The earliest theories concerning the scattering of bare charged particles by atoms were wholly classical and included the theories of Thomson(1) and Bohr(2) for β^-- and α-particle excitation of atoms and of Thomas(3) for re-arrangement of electron-capture collisions. The simplest quantum-mechanical descriptions of these processes were given very soon after the invention of quantum mechanics by Bethe(4), Oppenheimer(5) and Brinkman and Kramers(6) and were among the first quantum treatments of scattering processes. All of these theories discussed the impact of a bare nucleus and employed first-order perturbation theory in the interaction of the incident particle with the initially bound atomic electron. Not long afterwards, Massey and Smith(7,8) presented a non-perturbative approach, applicable to slow collisions and called the "perturbed stationary states" approximation. This approximation adopts the viewpoint that the collision is almost adiabatic so that electrons occupy molecular electronic states during the collision; transitions between such states are then effected by the motion itself, i.e. by the slight 'non-adiabaticity' of the collision. Although the original formulation contained several shortcomings, this simple idea has had far-reaching effects in the field of atomic and molecular collisions and, incidentally, has even found its way lately into descriptions of nuclear inelastic collisions.

After the initial activity of the early 1930's atomic collision research tended to take a back-seat to its nuclear counterpart in the 1940's and 1950's. The theory too consolidated only slowly. In 1948 Bohr(9) published his classic monograph on the penetration of charged particles in matter. Apart from a continuing undercurrent of theoretical work from the Massey school in the U.K. and a review of inner-shell ionisation by bare particle impact by Merzbacher and Lewis in 1958(10) little of importance was published. However, the 1960's saw a quickening of interest in atomic collisions and, in particular, in new experiments(11) on heavy atom-atom collisions where both collision partners carry electrons into the collision, where there is a deep interpenetration of the inner electron shells and where many modes of fragmentation of the atom-atom system (final channels) are available. (A few such experiments had been conducted in the 1930's(12) but had been subsequently forgotten.) This period coincided with the interest of T.P. Division in atom-atom collisions.

Although we shall mostly describe the contributions of the members of the Atomic and Molecular Physics Group under the leadership of Dr. John Tait in the period 1960-1979, it is not immodest to say that much of this work was highly innovative and often parallelled or led similar development of the theory in other laboratories. Hence the story of the work of T.P. Division presents us with a reasonably accurate overview of the recent development of interatomic collision theory. The contribution falls into two main parts. Up to 1970, interest concentrated on the lighter collision systems involving protons, α-particles and the helium and hydrogen atoms at collision energies in the range 1 keV - 1 MeV. After 1970 interest broadened to include the collisions of heavy atoms, such as the O^+-Ne collision system which was used as a prototype for the development of scaling laws for inner-shell excitation in any heavy-ion collision. At this time a comprehensive theory of photon emission accompanying a heavy-ion collision was also developed.

3. The Born Expansion and Beyond

It is fortunate that, largely as a result of the small ratio (m/M) of the electron mass to the nucleon mass, the nuclear motion can often be described accurately by a given classical trajectory of impact parameter $\underline{b}$ and initial collision velocity $\underline{v}$. Then the problem reduces to the calculation of the scattering wavefunction Ψ(r,t) of electrons moving in the time-dependent potential of their mutual interaction and that of the two nuclei. Hence a solution of the electronic Schrödinger equation

$$[H(t) - \frac{i}{\hbar}\frac{\partial}{\partial t}]\Psi(\underline{r},t) = 0 \qquad (1)$$

is sought, where H is the total Hamiltonian of the atom-atom system.

Although this classical trajectory approximation is not always admissible it will be used here for the sake of simplicity. The transition amplitude to some particular final state Φ_f is obtained as

$$f(\underline{b},\underline{v}) = \underset{t\to\infty}{\mathrm{Lt}} \langle\Phi_f(t)|\Psi^+(t)\rangle \tag{2}$$

where $\Psi^+(t)$ is that solution of eqn.(1) which propagates forward in time from an initial state $\Phi_i(t)$. The transition cross-section is obtained by integration over all impact parameters $\underline{b}$, i.e.

$$\sigma(\underline{v}) = \int d\underline{b}\,|f(\underline{b},\underline{v})|^2 \; . \tag{3}$$

The amplitude f can also be written as

$$f = \frac{i}{\hbar} \int_{-\infty}^{\infty} \langle\Phi_f(t)|V_f(t)|\Psi^+(t)\rangle \; dt \tag{4}$$

where $V_f(t)$ is that part of the total Hamiltonian which is not diagonalised in the final channel. In the earliest work the simplest first Born approximation was made by replacing the exact scattering wavefunction $\Psi^+(t)$ by the incident channel wavefunction $\Phi_i(t)$. For excitation and some types of ionisation this approximation gives the leading term in the total cross-section when the collision velocity far exceeds the electron orbital velocities. For slower collisions or for aspects of differential cross-sections this is not so and the first Born approximation fails. For re-arrangement collisions (electron capture) it does not even provide the leading term in fast collisions and the convergence of the Born series is in doubt. For this reason the early theories of atom-atom scattering are of very limited applicability.

In a series of papers Cheshire and co-workers put forward several ingenious methods to go beyond perturbation theory in the solution of the collision problem for light systems. The aim is to include distortion of the initial and final states as a result of intimate interaction with the Coulomb fields of the nuclei. Cheshire adopted a variety of methods to do this, depending upon the particular inelastic event under consideration. In most cases the simplest system $H^+ + H(1s)$ was considered. Then the three-body collision problem reduces to a one-body problem in the classical trajectory approximation and for hydrogen the <u>exact</u> undistorted wavefunctions are known. Cheshire's first method(13) was to expand both the wavefunction $\Psi^+(t)$ and the Coulomb operator $V_f(t)$ in partial waves about the target nucleus as origin. This leads to a set of coupled equations for the amplitude of each partial wave which can be solved numerically to give the <u>exact</u> scattering amplitudes. As the partial wave sum must be truncated to small angular momentum values to make the problem tractable, the method is most suited to excitation or ionisation of electrons into target states of small angular momentum. Since it uses a target-based expansion it is not suitable for charge exchange.

To handle the problem of charge exchange and the strong distortion required to transfer an electron from a bound state around one nucleus to a bound state around the other, Cheshire and co-workers suggested several methods(14) of which the most useful has proved to be the 'continuum distorted-wave' method.

The continuum distorted-wave method is an example of the general class of distorted-wave approximations much used in nuclear physics. The final state, $\Phi_f(t)$, satisfies

$$H_f \, \Phi_f = i\hbar \frac{\partial}{\partial t} \Phi_f \tag{5}$$

where H_f is the final channel Hamiltonian. The undistorted state Φ_f is replaced by a distorted function χ_f which satisfies

$$(H_f + U_f)X_f = i\hbar \frac{\partial}{\partial t} X_f \tag{6}$$

when the transition amplitude becomes, instead of (4)

$$f = \frac{i}{\hbar} \int_{-\infty}^{\infty} \langle X_f(t) | (V_f - U_f) | \Psi^+(t) \rangle \, dt \tag{7}$$

The distorting potential U_f is arbitrary but is chosen to give a high overlap of the states χ_f with the continuum intermediate states involved in charge transfer. Higher order distorted-wave approximations are obtained by approximating the exact wave function Ψ successively by the initial wave-function Φ_i and its distorted version χ_i. This method was applied successfully by Cheshire to proton-hydrogen collisions and recently has found wide application in the asymmetric collision systems(15) of interest in fusion plasmas.

Charge transfer under Coulomb forces is peculiar in that at high velocity the second Born term gives a larger contribution to the cross-section than does the first Born term. Cheshire has shown that his distorted-wave method gives the correct high-velocity limit of the second Born approximation. Briggs(16) has considered an alternative distorted-wave approximation, misleadingly called the impulse approximation, and demonstrated that this approximation too provides the correct second Born limit, contrary to what was previously believed(17).

It is interesting that the theoretical prediction of the importance of the second Born term has not to date been confirmed experimentally. The dominant second-order term has been shown(18) to correspond exactly in the limit of large quantum numbers to a classical double-scattering capture mechanism considered by Thomas as long ago as 1927(3). For many-electron targets Thomas showed that this double-scattering can lead to the capture of one electron and the simultaneous ejection of another electron with a fixed momentum. Only recently have Briggs and Taulbjerg(19) (a frequent visitor to the Division from Aarhus, Denmark) shown that a quantum-mechanical treatment of this process also leads to a characteristic ejected electron spectrum whose observation may provide a convenient experimental signature of the second Born term. In the same vein, Dubé(20), another member of T.P. Division, has generalized the second Born theory to include transfer between atomic states of arbitrary quantum number and suggested that the observation of the polarisation of photons emitted upon the decay of excited electron states formed by capture could also lead to identification of the asymptotically dominant double-scattering process described by the second Born approximation.

4. Close-Coupling Expansions

Such studies of the fundamental behaviour of electron re-arrangement collisions are of rather academic interest since they involve high-velocity expansions, although with hydrogenic wave-functions analytic results may be obtained. However, to describe collisions at velocities of experimental interest, resort is made to the greater flexibility of numerical methods. The most useful non-perturbative approach, which has been developed extensively in T.P. Division over the years, (also for electron scattering, see the article by P.G. Burke in this volume), is the method of close-coupling or eigenfunction expansions. The basic idea is simple. The state vector $|\Psi\rangle$ representing the exact scattering wave-function is approximated by a linear combination of known functions $|\psi_n\rangle$, i.e.

$$|\Psi\rangle = \sum_n a_n|\psi_n\rangle = \underline{a}.|\psi\rangle \ , \qquad (8)$$

where $\underline{a}$ and $|\psi\rangle$ are column vectors of finite dimension N. The best approximation to the exact state $|\Psi\rangle$ in a variational sense is given by the vector $\underline{a}$ which satisfies the set of coupled differential equations,

$$i\hbar\ \underline{\underline{S}}.\left(\frac{d\underline{a}}{dt}\right) = \underline{\underline{M}}.\underline{a} \ , \qquad (9)$$

where S is the overlap matrix with elements $S_{mn} = \langle\psi\,|\,\psi\,\rangle$ and M is the coupling matrix with elements

$$M_{mn} = \langle\psi_m|H - i\hbar\,\frac{\partial}{\partial t}\,|\psi_n\rangle \ .$$

The numerical solution of the N-dimensional problem (9) subject to appropriate boundary conditions allows the transition amplitude to any final state to be calculated via equations (8) and (2). Although simple in principle, in practice the necessity to preserve the Galilean invariance of the time-dependent Schrödinger equation under the frame transformations which occur in re-arrangement collisons imposes a severe restriction on the form of the basis states ψ_n , which leads to considerable calculational difficulty.

Cheshire, Gallaher and Taylor(21) performed probably the most sophisticated calculations of this type using basis functions of bound atomic character augmented by "pseudo-states" designed to represent the continuum adequately. When applied to proton-hydrogen atom collisions involving excitation and charge transfer to low-lying states a close agreement with the experimental results was obtained and the considerable structure in total and differential cross-sections explained.

When the collision velocity is less than electronic orbital velocities, the formation of a collision complex signals the necessity to use basis functions which are eigenstates of the temporary molecule, a feature first emphasised in this context by Fano and Lichten(22). Then the solution of eqn. (9) corresponds closely to the perturbed stationary states method of Massey and Smith(7). In order to retain the calculational simplicity of atomic states, Cheshire(23) proposed to approximate the molecular states by atomic states defined with a time-varying nuclear charge representing molecule formation. This variable charge is decided by a subsidiary variational condition (although the variational problem is now non-linear). This method has recently been generalized(24) to allow a complex effective charge to represent loss into unobserved inelastic channels. Explicit calculations on the proton-hydrogen atom collision at low velocity using molecular basis states have also been performed and these combine smoothly with those of Cheshire et al as the collision velocity is raised.

The success of close-coupling methods in simple systems is hard to repeat when both colliding atoms carry several electrons. This is due to the complexity of the many-body problem, particularly regarding the electron-electron interaction. However, as a result of pressure to understand the inelastic processes involved in heavy-atom collisions, Briggs, Macek, Taulbjerg and Vaaben have, in the period since 1971, developed methods(25) which, at least for inner-shells, allow the calculation of cross-sections to be made in a close-coupled basis of Hartree-Fock molecular states. The development of scaling laws for inner-shell excitation(26) was an important advance allowing rapid and universal application of these results. To-day, with the exception that the relativistic time-dependent Dirac equation rather than the Schrödinger equation (1) must be used, the same method is used to discuss inner-shell vacancy formation in, for example 1 GeV U^+ - U collisions(27) - remarkable energy at which to do atomic physics! In such collisions, the binding energies of electrons in temporary molecular levels can exceed the electron rest mass. Then the same methods can also be used to discuss positron states and positron emission(28).

5. Continuum X-ray Emission

Before 1971, only photons of a discrete frequency corresponding to the decay of levels in isolated atomic or ionic fragments after the collision had been observed in single ion-atom collision experiments. However, in a variety of experiments since then(29), both at low and high velocity, new types of continuum emission, corresponding to radiative decay processes in the collision complex, have been identified. Some of the earliest experiments of this type were performed by Cairns and Marwick of Metallurgy Division. Briggs and Dettmann(30) formulated a unified theory of photon emission in heavy-ion collisions incorporating both characteristic and continuum frequency photons. The amplitude for emission of a photon with frequency and polarisation vector $\underline{e}$ is written

$$f_\lambda(\omega) = (2\pi\sqrt{\omega})^{-1} \int_{-\infty}^{\infty} dt <\Psi_f^-(t)|\underline{e}_\lambda.\underline{p}|\Psi_i^+(t)>e^{i\omega t} \qquad (10)$$

where p is the total electron momentum and $\Psi_i{}^+$, $\Psi_f{}^-$ are forward and backward time-propagating solutions of eqn. (1). It is apparent from equation (10) that observation of photon distributions provides, in a Fourier analysis sense, information on the time development of the collision complex. By contrast, observation of final channels only, via eqn. (2), provides only information on the $t\to\infty$ behaviour of the scattering wavefunction. Hence the photon-fragment co-incidence experiments performed nowadays provide a far more sensitive test of collision theories. As a consequence, the theory of Briggs and Dettmann has been suitably generalized(31) to discuss the polarisation and angular distribution characteristics of emitted radiation. An interesting result of this work was the prediction(32) that photon emission provides the asymptotically most probable mechanism of electron capture, a result of some importance for the studies of asymptotic capture cross-sections mentioned previously.

6. Technical Applications

Although in the foregoing we have emphasized the more fundamental nature of the development of atomic collision theory over the past twenty years or so, many detailed applications of the methods have been made by members of the Atomic and Molecular Physics Group in this and related fields. Thus, Cheshire and Poate(33) and Briggs and Pathak(34) discussed the stopping power of crystalline solids for channelled heavy-ions. Drepper and Briggs(35) and more lately Day(36) are developing the theory of the momentum distribution of ejected electrons (the delta-rays of radiation physics) when atoms collide. Vaaben and Briggs(37) and Greenland(38) have used the molecular states expansion method to discuss the population inversion resulting from charge transfer in such asymmetric systems as C^{6+}+H(1s), which are of importance to the problem of impurities in fusion plasmas. Hodgkinson and Briggs(39) used similar methods to describe symmetric charge transfer in the outer shells of colliding heavy atoms at thermal velocities. The same authors have contributed to the theory of fragmentation of atoms and molecules under laser irradiation(40). This is the inverse of the collisional emission problem considered above and, since the light source is strong, it requires an appropriate generalization of the perturbative interaction of the collision complex with the radiation field embodied by eqn. (10). Tait, Taylor, Faisal, Burke and Scrutton(41) considered the recombination of oxygen atoms in collision with CO molecules as an aid to the understanding of graphite corrosion processes.

So it is seen that the work of Atomic and Molecular Physics Group in the field of atomic collisions over the past twenty-five years has involved a marriage of fundamental development and technical application, the hallmark of the operations of Theoretical Physics Division.

1. J.J. Thomson, Phil. Mag. 23, 449 (1912).

2. N. Bohr, Phil. Mag. 25, 10 (1913); 30, 581 (1915).

3. L.H. Thomas, Proc. Roy. Soc. 114, 561 (1927).

4. H.A. Bethe, Ann. der Phys. 5, 325 (1930).

5. J.R. Oppenheimer, Phys. Rev. 31, 349 (1928).

6. H.C. Brinkman and H.A. Kramers, Proc. Acad. Sci. (Amsterdam), 33, 973 (1930).

7. H.S.W. Massey and R.A. Smith, Proc. Roy. Soc. A114, 188 (1934).

8. N.F. Mott and H.S.W. Massey in The Theory of Atomic Collisions, 1st Edn. (Oxford University Press, 1933).

9. N. Bohr, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. No. 18 (1948).

10. E. Merzbacher and H.W. Lewis in Handbuch der Physik, Ed. S. Flügge, Vol. 34, p.166 (Springer-Verlag: Berlin, 1958).

11. V.V. Afrosimov, Yu S. Gordeev, M.M. Panov and N.V. Federenko, Sov. Phys. - Tech. Phys. 9, 1248, 1256, 1265 (1965) and E. Everhart and Q.C. Kessel, Phys. Rev. Lett. 14, 247 (1965).

12. O. Beeck, Ann. der Phys. 6, 1001 (1930) and W.M. Coates, Phys. Rev. 46, 542 (1934).

13. I.M. Cheshire and E.C. Sullivan, Phys. Rev. 160, 4 (1967).

14. I.M. Cheshire, Phys. Rev. A138, 992 (1965) and Proc. Phys. Soc. 84, 89 (1964).

15. R. Shakeshaft, J. Phys. B7, 1734 (1974). and D.Z. Belkic, J. Phys. B10, 3491 (1977).

16. J.S. Briggs, J. Phys. B10, 3075 (1977).

17. M.R.C. McDowell and J.P. Coleman in Introduction to the Theory of Ion-Atom Collisions, (N. Holland: Amsterdam, 1970) p.418.

18. L. Spruch, Phys. Rev. A18, 2016 (1978).

19. J.S. Briggs and K. Taulbjerg, J. Phys. B12, 2565 (1979).

20. J.S. Briggs and L. Dubé, J. Phys. B. (in press).

21. I.M. Cheshire, D.F. Gallaher and A.J. Taylor, J. Phys. B3, 813 (1970).

22. U. Fano and W. Lichten, Phys. Rev. Letts. 14, 627 (1965).

23. I.M. Cheshire, J. Phys. B1, 428 (1968).

24. M. Kleber, to appear in J. Phys. B.

25. J.S. Briggs, Rep. Prog. Phys. 39, 217 (1976).

26. M. Kleber, J. Phys. B11, 1069 (1978).

27. B. Müller, G. Soff and W. Greiner, Zeits. für Phys. A285, 27 (1978).

28. W. Betz, G. Heiligenthal, J. Reinhardt, R.K. Smith and W. Greiner in The Physics of Electronic and Atomic Collisions, eds. J.S. Risley and R. Geballe, (Univ. of Washington Press: Seattle, 1975) p.531.

29. W.E. Meyerhof and K. Taulbjerg, Ann. Rev. Nuc. Sci. 27, 279 (1977).

30. J.S. Briggs and K. Dettmann, J. Phys. B10, 1113 (1977).

31. J.S. Briggs, J.H. Macek and K. Taulbjerg, J. Phys. B12, 1457 (1979).

32. J.S. Briggs and K. Dettmann, Phys. Rev. Letts. 33, 1123 (1974).

33. I.M. Cheshire and J.M. Poate in Atomic Collision Phenomena in Solids, eds. D.W. Palmer, M.W. Thompson and P.D. Townsend, p.351 (1970).

34. J.S. Briggs and A.P. Pathak, J. Phys. C.6, L173 (1973).

35. F. Drepper and J.S. Briggs, J. Phys. B9, 2063 (1976).

36. M. Day, J. Phys. B. (to be published).

37. J. Vaaben and J.S. Briggs, J. Phys. B10, L521 (1977).

38. P.T. Greenland, J. Phys. B11, L191 (1978).

39. D.P. Hodgkinson and J.S. Briggs, J. Phys. B9, 255 (1976).

40. D.P. Hodgkinson and J.S. Briggs, J. Phys. B10, 2583 (1977).

41. J.H. Tait, A.J. Taylor, F.H.M. Faisal, P.G. Burke and D. Scrutton, VIth ICPEAC, abstracts p.630 (M.I.T. Press: Cambridge, Mass., 1969) and J. Phys. B2, 1155 (1969).

Chapter VII

DEVELOPMENTS IN PLASMA PHYSICS AND CONTROLLED FUSION

William B. Thompson

1. Introduction

In the twenty-five years since the founding of the Authority, the pursuit of thermonuclear fusion has expanded from a preoccupation of a small group of eccentrics to a major research industry employing thousands and spending hundreds of millions of pounds. It would be nice to report that as spending has exceeded the initial estimates, progress has lived up to the expectations of that early group. In spite of the thousands and millions, however, there is still no evidence for the release of significant amounts of energy in any controlled fusion experiment. Those fusion reactions which have been produced have made themselves known only through countable fluxes of neutrons, i.e. via very sensitive detectors.

Although the practical production of fusion power remains a goal glimpsed in the distance (perhaps it is even a mirage), in its pursuit a good deal has been learned about the physics of plasmas, which has had some impact outside the fusion field. Indeed, astrophysicists now find plasmas everywhere, from the earth's ionosphere to quasars, perhaps the most distant observable objects; the only non-plasma place being the solid earth, and even here solid state physicists have invoked plasmas to explain such processes as the anomalous transmission of radio waves through metal films and fluctuations in the high current conductivity of semiconductors.

It is, of course, not possible to give an adequate account of the last two decades of developments in plasma physics, but let me outline a few.

2. Oscillations, Waves and Instabilities

Perhaps the simplest class of problem is that of small oscillations about a given equilibrium, since these can be treated as linear. Simple fluid models of uniform systems predict high-frequency plasma oscillations, in which electrons and ions move in opposite directions, and modified sound waves in which they move together. In a uniform magnetic field the complex phenomena predicted by the Appleton-Hartree theory of ionospheric propagation appear even if thermal effects are ignored(1). A major advance was made by Vlasov(2). He observed that on the relevant time scale collisions were unimportant, and that rather than behaving as a gas with a local equation of state, the electrons move largely independently, and should be described by a distribution function f and its disturbance $\delta f(x,v,t)$ which satisfies a Boltzmann equation with no collisions. This led to a dispersion relation in the form*

$$k^2 = \omega_p^2 \int \frac{\underline{k}.\frac{\partial f}{\partial \underline{v}}}{(\omega + \underline{k}.\underline{v})} d^3v \quad . \tag{1}$$

This describes not only the high frequency oscillations, but at low frequencies the Debye screening of a charge in a plasma. If, in addition to the electrons, the contribution of the ions is allowed for, then Vlasov's dispersion relation will also describe the plasma analogue of normal sound waves.

* The symbols used in this and the subsequent equations are mostly 'standard' but their definitions are collected together at the end of this article.

Landau(3) criticized the work of Vlasov, observing that if the solution is to be causal, the pole in the integral must be retained. This has the rather surprising consequence of inducing an imaginary term

$$i\pi\,\omega_p^2 \int \underline{k}.\frac{\partial f}{\partial \underline{v}}\,\delta(\omega + \underline{k}.\underline{v})\,d^3v \quad , \tag{2}$$

into the dispersion relation and hence to the collisionless damping of oscillations. Such damping is usually associated with an increase in entropy, but, in the Vlasov equation, the distribution function f and hence the entropy

$$S = \int f \ln f d^3v$$

are constants. During the early 50's there was some theoretical controversy on the subject of Landau damping, which was of great importance in the developing theory of plasma stability.

A major experimental advance was the experimental demonstration of this phenomenon by Malmberg and Wharton(4). The experiment required the transmission of Langmuir waves on a neutralized electron beam, which travelled parallel to a strong magnetic field, so that the disturbed motion was one-dimensional. The distribution function and the perturbed electric field were measured by probes. The damping of waves was found to depend in the correct way on the distribution function. Moreover, the electron distribution could be 'clipped' so that both f and $\frac{\partial f}{\partial v}$ were zero at the phase velocity of the wave, whereupon the damping disappeared.

The importance of this development lies in the central role of Landau damping in plasma stability theory. Although some types of instability can be more or less described by representing the plasma as an ideal conducting fluid, and then considering the sign of variations in the potential energy integrated over the plasma volume

$$\text{P.E.} = \int\left(\frac{p}{(\gamma-1)} + \frac{B^2}{8\pi}\right)d^3x \quad , \tag{3}$$

or by extending the fluid model to include finite resistivity, there are many other modes of instability that arise because the Landau 'damping' term has the wrong sign for damping. The simplest of these are velocity-space instabilities in spatially uniform systems; for example, plasma waves, or electromagnetic waves driven by the free energy in the non-Maxwellian distribution(6). In complex magnetic geometries it is often necessary to use some approximation to get at the distribution function, often an expansion in powers of the gyro-radius which uses the adiabatic invariants of particle motion(7).

Almost any source of free energy can be trapped to drive some instability. However, the Landau modes usually abstract energy from only a small class of particle and although hard to stabilize, they do not usually disrupt an equilibrium violently, but lead instead to an anomalous diffusion which may or may not be tolerable.

3. Plasma Turbulence

A further important development has been in the non-linear theory of plasma oscillations. There have been four developments here, namely (i) quasilinear theory, (ii) the theory of parametric instability, (iii) the wave kinetic theory and (iv) the single wave theory. The first of these starts by constructing the Boltzmann analogue of Reynolds stresses. Thus if f and E are the fluctuations in the distribution function and the field, then the mean value of the distribution function f satisfies

$$\frac{\partial \bar{f}}{\partial t} + \underline{v}.\underline{\nabla}\bar{f} + <\frac{e}{m}\underline{E}.\frac{\partial}{\partial \underline{v}}\delta f> = 0 \quad . \tag{4}$$

If the system is unstable, the electric field amplitude develops as

$$\frac{\partial |E|^2}{\partial t} = \omega_p^2 \int \underline{v}.\frac{\partial f}{\partial \underline{v}}\,\delta(\omega + \underline{k}.\underline{v})\; d^3v|E|^2 \tag{5}$$

while

$$\frac{e}{m} <\underline{E}.\frac{\partial}{\partial \underline{v}}\,\partial f> = \frac{e^2}{m^2}\frac{\partial}{\partial \underline{v}} \cdot \int |E|^2\,\delta(\omega + \underline{k}.\underline{v})\,\frac{\partial f_0}{\partial \underline{v}}\,d\omega d^3\underline{k} \quad . \tag{6}$$

This pair of equations gives a description of the distortion in the average value of the distribution function, and the switching off of a velocity space instability(8). Under some circumstances the description seems adequate. An important and successful application of the quasi-linear theory has been unravelling the loss process in mirror machines where the instabilities are driven by the loss cone in velocity space.

If non-linear effects are important, then waves can scatter off each other. An important role is played by the plasma number $n_k = \varepsilon_k/\omega_k$ which is conserved on interaction, and is determined by the wave analogue of the Boltzmann equation

$$\frac{\partial n_k}{\partial t} + \underline{\nabla}.\underline{v}_g\, n_k - 2\gamma n_k = \int \delta(\omega_k - \omega_{k'} - \omega_{k''})\; V_{k,k',k''}(n_k n_{k'} - n_{k'} n_{k''}) d^3k' d^3k'' \tag{7}$$

where v_g is the group velocity, γ is the linear damping coefficient and V is a scattering matrix. The study of this equation has formed a considerable industry(9).

If a single intense pump wave is applied to a plasma, this may couple a pair of plasma modes together - the classical example being the coupling of a Langmuir wave to an ion acoustic mode by a transverse wave close to the plasma frequency. The coupled wave kinetic equations may then admit exponentially-growing unstable solutions which can lead to greatly enhanced power absorption(10). These processes are of great importance in the anomalous back-scattering of intense laser beams. For example, an incoming and a reflected electromagnetic wave may combine to produce either a Langmuir plasma wave, as in induced Raman scattering, or an ion acoustic wave, as in induced Brillouin scattering - processes which are predicted to be of importance when the spatial scale length is great compared to the optical wave length. In most present day experiments these are less important than the resonance absorption that occurs at non-normal incidence, where the transverse electric field couples with the density gradient to produce a rapidly amplifying longitudinal wave. Near the critical surface defined by the equation $\omega = \omega_p$, where the electrostatic wave is resonant, field strengths can become so large that the ponderomotive force, $\nabla(E^2/8\pi)$, exceeds the plasma pressure with the result that the wave digs a hole, the so-called 'caviton', in the plasma. These 'cavitons' are typical of a class of coherent non-linear structures which have attracted a good deal of attention(11). Unstable Langmuir waves may grow to such a value that cavitons are formed. If less free energy is available, growing Langmuir waves saturate by trapping resonant particles in a potential well(12). Non-linear ion acoustic modes may take the form of solitons, persistent solutions to a non-linear wave equation that can survive collisions(13). Another non-linear structure is the more or less persistent double charge layer, a region of locally high electric field.

A major problem which is currently fashionable is that of the relation between such coherent non-linear structures and the chaotic behaviour characteristic of turbulence or of statistical mechanical systems generally. It is here that plasma theory comes close to exciting recent developments in classical mechanics(14).

4. Basic Kinetic Theory

The classical work of Chapman and Cowling(15) contained a development of the kinetic theory of ionized gases which included the effect of magnetic fields strong enough that the cyclotron frequency Ω was comparable to the collision frequency ν. Although this work retains its value, the treatment of the Boltzmann collision integral was rather arbitrary. Thus, in its primitive form, the collision integral diverges because of the long range of the Coulomb force, so that many-particle rather than two-particle encounters determine the evolution of the distribution function. Landau showed,(16) as Jeans had done before him in a discussion of stellar collisions(17), how the Boltzmann integral, for small collisions, led to a Fokker-Planck equation, viz.

$$\frac{\partial f}{\partial t} + \underline{v}.\underline{\nabla} f = \frac{\partial}{\partial \underline{v}}.J(f,f) \tag{8}$$

where

$$J = \frac{e^4}{m^2} \ell n\Lambda \int \frac{d^3\bar{v}}{g} (1 - \hat{g}\hat{g})\left(\frac{\partial}{\partial\bar{v}} - \frac{\partial}{\partial v}\right) f(\bar{v})f(v)$$

and

$$\underline{g} = \underline{v} - \underline{\bar{v}}$$

although again an arbitrary cutoff was needed, and hidden in Λ.

By using a modification of the theory of liquids, as developed by Bogoliubov, Born, Green, Kirkwood, Yvon and others, in which one starts not from the Boltzmann function f(x) but from the Liouville function, $F(x_1..x_N,v_1..v_N)$, it has proved possible to develop a fairly consistent kinetic theory of the plasma, the Balescu-Lenard equation(18). This can be more easily developed by analogy with the quasi-linear theory, except that the fluctuating fields have their origin not in instabilities but in the discreteness of charged particles(19). The diffusion current J then may be written

$$J = \int d^3\bar{v} \int d^3k d\omega \frac{\underline{kk}\ \delta(\omega + \underline{k}.\underline{v})}{|k^2\ \varepsilon(\omega,k)|^2} \left(\frac{\partial}{\partial\bar{v}} - \frac{\partial}{\partial v}\right) f(\bar{v})f(v) \quad . \qquad (9)$$

If only the large k behaviour of ε is taken into account this reduces to Landau's form, but without the use of an arbitrary long range cut off.

A more subtle consequence of this development is that the electron correlation function may be calculated; furthermore, since this determines the radiation scattering properties of the plasma, it may be measured. Attention was drawn to this problem when the Doppler shift of back-scattered radiation was used by Bowles(20) to estimate the electron temperature in the ionosphere, with ridiculously low results. Dougherty and Farley(21) showed that if the change in wave number k was less than the Debye wave number, k_D, then scattering was dominated not by fluctuations in the free electrons but by the correlations induced by the Debye shielding clouds around ions. Ever since these measurements, the scatter of radiation has proved a most important method of investigating plasma.

A further important development was the long mean-free-path version of kinetic theory. In a magnetically confined plasma it is reasonable to think of kinetic processes in which the collision frequency is not the highest but the lowest frequency in the problem; and then to discuss kinetic theory backwards, as it were, putting in particle dynamics and geometrical effects before collisions. An interesting result of this is the 'banana' diffusion in Tokomaks, in which collisions induce transitions between drift surfaces, which can be separated by much more than a Larmor radius. As a result, diffusion is determined not by the classical $\nu r_L{}^2$, but exceeds this by a ratio of order B_ϕ/B_θ which is usually large(22). Although the neo-classical theory comes close to a description of the observed diffusion in Tokomak systems, there remain discrepancies arising from instabilities. In the less dense regions of the plasma, a current-driven ion wave leads to quasi-linear diffusion, while near the centre an unstable tearing mode disrupts the flux surfaces, forming magnetic islands and permitting a radial transport along magnetic field lines that imitates diffusion. In spite of some succeses, plasma kinetic theory is as yet incomplete. Long range correlation and wave-induced transport are important, and the role of non-linear coherent structures is still unclear.

5. Developments in Fusion

Great advances have been made possible in relating theory to experiment by the spectacular development of computation. Even in the best designed experiments, the detailed comparison with theory usually calls for numerical solutions. By use of such methods a body of dependable understanding is slowly being developed. Meanwhile, in fusion experiments, computation has permitted the analysis and design of fairly complex equilibrium configurations. In the analysis of stability, the computer again plays an essential role. An important recent development has been programs which can assess ideal stability by explicitly computing the change in fluid potential for an arbitrary displacement(23). In the study of resistive modes, numerical solutions to the eigenvalue problem permit the analysis of realistic geometries. These studies have shown, for example, that Tokomaks of non-circular cross-section can have improved stability. More generally, they have had a considerable impact in the experimental program. In the analysis of diffusion, particularly of anomalous diffusion, computation has been essential in relating theory to experiment.

In kinetic theory, particle simulation codes have helped our understanding of plasma behaviour, particularly in showing how instabilities develop.

Even in the theory of linear wave propagation computation has permitted detailied studies of absorption and has suggested that heating can be localized on particular internal surfaces. This has been demonstrated(24) in the 'Bumpy Torus' system, and it raises the possibility of controlling the temperature and pressure profile. This might make it possible to preserve a shear stabilized Z-pinch in stable equilibrium, a much happier candidate for a practical thermonuclear reactor than is a Tokomak.

Meanwhile, new approaches to fusion, in particular the possibility of producing micro-explosions by compressing and heating deuterium-tritium pellets by external pulsed power sources, open new challenges to plasma theory[25]. What happens as a real plasma interacts with intense radiation? What the the properties of strongly interacting and partially degenerate plasmas?

Although this approach initially depended on lasers as drivers, other possible drivers are being studied, especially ion beams. Currents of millions of amperes at energies of a few MeV might be produced by the steadily developing pulsed power technology which is based on Marx generators and water storage lines; while thousands of amperes at a few GeV of heavy ions might be produced by developments of linear accelerators and storage rings[26]. Major plasma problems are concerned with the propagation and focussing of the ion beams, as well as with target interaction.

Our knowledge of plasma physics grows slowly and painfully, for, in spite of their importance, basic studies do not use a significant fraction of fusion budgets. Nonetheless, even though a practical fusion reactor is still around the corner, significant progress has been made and we can now approach problems on a fairly sophisticated level - although we cannot yet claim to have anything like an adequate understanding of the subject.

1. J.A. Ratcliffe, "The Magnetoionic Theory" (Cambridge Univ. Press, 1959).

2. A. Vlasov, J.E.T.P. 8, 291 (1938).

3. L. Landau, J. Phys. U.S.S.R. 10, 25 (1946).

4. J. Malmberg and C.B. Wharton, Phys. Rev. Lett. 13, 184 (1964).

5. I.B. Bernstein, E.A. Frieman, M.D. Kruskal and R.M. Kulsrud, Proc. Roy. Soc. A 244, 17 (1958).

6. See, e.g., N.A. Krall and M.N. Rosenbluth, Phys. Fluids 6, 254 (1963).

7. See, e.g., W.B. Thompson, "An Introduction to Plasma Physics" (Pergamon Press, Oxford, 1962).

8. W.E. Drummond and D. Pines, Nuclear Fusion Suppl. 3, 1049 (1962).

9. See, e.g., E.G. Harris, Adv. Plasma Phys. 3, 157 (1969).

10. See, e.g., D.F. Dubois and M.V. Goldman, Phys. Fluids 8, 1404 (1965).

11. V.E. Zakharov, J.E.T.P. 35, 908 (1972).

12. T.M. O'Neil, Phys. Fluids 8, 2255 (1965).

13. N.J. Zabusky and M.D. Kruskal, Phys. Rev. Lett. 15, 240 (1965).

14. See, e.g., M.V. Berry in "Topics in Non-Linear Dynamics", ed. S. Jorna, p. 16, American Inst. Phys., New York (1978).

15. S. Chapman and T.E. Cowling, "The Mathematical Theory of Non-Uniform Gases", 2nd edition (Cambridge Univ. Press, 1952).

16. L. Landau, J.E.T.P. 7, 203 (1937).

17. J.H. Jeans, "Astronomy and Cosmogony" (Cambridge Univ. Press, 1929).

18. See, e.g., A. Lenard, Ann. Phys. (N.Y.) 3, 390 (1960).

19. See, e.g., W.B. Thompson and J. Hubbard, Rev. Mod. Phys. 32, 714 (1964).

20. K.L. Bowles, Phys. Rev. Letts. 1, 454 (1958).

21. J.P. Dougherty and D.T. Farley, Proc. Roy. Soc. A 259, 79 (1960).

22. See, e.g., A.H. Glasser and W.B. Thompson, Phys. Fluids 16, 95 (1973).

23. R. Moore, Bull. Am. Phys. Soc. 24, 1044 (1978).

24. R.A. Dandl and G.E. Guest, "The ELMO Bumpy Torus", to appear in "Fusion", ed. E. Teller (Academic Press, New York, 1980).

25. See, e.g., K.A. Brueckner and S. Jorna, Rev. Mod. Phys. 46, 325 (1974).

26. See, e.g., W.B. Hermannsfeldt, "The Development of Heavy Ion Acceleration for Inertially Confined Fusion", Lawrence Berkeley Lab. Report, LBL 9332.

List of Symbols

$\underline{B}$	Magnetic Induction
B_ϕ and B_θ	Angular components of magnetic induction in spherical co-ordinates.
c	Velocity of light.
e	Electronic charge.
$\underline{E}$	Electric field vector.
E_k	Energy in mode k.
f	Distribution function for electrons.
f_o	Equilibrium distribution function.
$\underline{g}$	Relative velocity vector, $\underline{g} = \underline{v} - \underline{\bar{v}}$.
$\hat{\underline{g}}$	$\hat{\underline{g}} = \underline{g}/\lvert g\rvert$.
$\underline{k}$	Wave-vector of a disturbance component.
k_B	Boltzmann's constant.
m	Electron mass.
n	Electron number density.
r_L	Larmor radius, $mcv_\perp/eB$.
p	Fluid pressure.
T	Absolute temperature.
$\underline{v}$	Electron velocity.
$v_\perp$	Tangential velocity of an electron round the field line.
$V_{kk'k''}$	Effective potential for mode interaction.
$\underline{x}$	Electron position vector.
γ	Ratio of specific heat at constant pressure to that at constant volume.
γ	Linear damping coefficient to plasma wave.
$\varepsilon(\omega,k)$	Dielectric function of the plasma.
λ_D	$(k_BT/4\pi ne^2)^{\frac{1}{2}}$ is the Debye screening length.

Λ	$1/4\pi n\lambda^3$, the Coulomb screening factor.
ν	Effective collision frequency.
ω	Frequency of a disturbance component.
ω_k	Frequency of mode k.
ω_p	$(4\pi ne^2/m)^{\frac{1}{2}}$, the plasma frequency.

Chapter VIII

THE PHYSICS OF MAGNETIC FUSION ENERGY

K. V. Roberts

This will be a personal account covering the period from April, 1956, when Kurchatov gave his famous lecture at Harwell and thus began the process of declassification of fusion research, until the present day. In 1956 I was at A.W.R.E. Aldermaston and did not join Harwell until 1959, but I was already very interested in the challenging problem of the controlled release of energy by magnetic confinement fusion. Even after 24 years it remains an elusive but appealing goal.

In order to release fusion energy from the light isotopes the material must be raised to a high temperature of at least several keV so that the positively-charged nuclei are energetic enough to overcome their mutual Coulomb repulsion, and it must be held together long enough against a natural tendency to fly apart at thermal velocity in order to release more than its initial thermal energy by nuclear reactions - the well-known Lawson criterion(1). At such temperatures, it will be a fully ionized gas, except perhaps for heavy impurity ions. The three known confinement methods are inertial (big bang, supernovae, weapons and laser fusion), gravitational (ordinary stars) and magnetic. The third method is by far the most difficult to achieve and to analyse, and while inertial and gravitational systems can be spherically symmetric, magnetic systems must inevitably have a complicated geometry. Magnetic confinement relies on the fact that, although the plasma is electrically neutral, the individual charged ions and electrons move in helical Larmor orbits and to a good approximation, remain attached to the field lines, behaving like small electrically-charged magnets with adiabatically-constant magnetic moment $\mu = mV_{\perp}^2/2B$ where $V_{\perp}$ is the perpendicular thermal velocity. Two main classes of magnetic confinement system exist, closed-line (topologically toroidal) and open-line (mirror). Confinement in open-line systems depends on the particles behaving as if they had potential energy μB so that they are reflected from mirror regions of high magnetic field B, at the ends of the apparatus. In practice both situations are complicated by several effects, viz. (i) changes in magnetic moment μ and cross-field diffusion due to collisions, (ii) drifts due to non-uniform magnetic or transverse electric fields, (iii) reflection due to longitudinal electric fields, (iv) the effect of turbulent electric and magnetic fields caused by instabilities, and (v) by the fact that closed-line systems possess internal mirror regions.

Although several thermonuclear reactions are known, at present one envisages only the use of the fastest reaction

$$D + T \rightarrow He^4 + n + 17.6\text{Mev} \quad .$$

Since tritium is mildly beta-active with a half-life of 12.6 y, it must be regenerated in a lithium blanket by the reaction

$$n + Li^6 \rightarrow T + He^4 + 4.8\text{Mev}$$
$$n + Li^7 \rightarrow T + He^4 + n - 2.5\text{Mev} \quad .$$

Other (n,2n) reactions may also contribute and the tritium breeding time can be quite short - of the order of months. The thermonuclear reaction rate is very sensitive to temperature, and to overcome the loss of energy due to bremsstrahlung, which varies as $T^{\frac{1}{2}}$, the plasma must be raised to the ideal ignition temperature of about 5 keV, or even higher if energy losses due to impurity radiation and thermal conduction are taken into account. The optimum operating temperature is about 15 keV for closed-line devices but higher for mirror machines. Practical magnetic field strengths determine a maximum plasma pressure and hence a maximum plasma density of about 10^{14}cm^{-3}, and this, together with the Lawson criterion

$$n\tau \gtrsim 3.10^{14} \text{ s cm}^{-3}$$

means that the containment time, τ, must be at least several seconds.

By 1958 all this was familiar and in addition to the considerable effort on the toroidal Z-pinch going on at Harwell and at A.E.I. Aldermaston, there was work on the linear Θ-pinch and the mirror at A.W.R.E. The simultaneous announcement in Nature(2) in January, 1958, of the work on ZETA and SCEPTRE and of research in the U.S.A. in January, 1958, was followed by the release of a flood of previously-classified information from the U.K., U.S. and U.S.S.R. at the Second Geneva Conference in September, 1958. Bill Thompson gave a series of lectures on controlled thermonuclear reactions (C.T.R.) which were later published as a book(3) while Walter Marshall lectured on kinetic theory(4). Roger Taylor lectured extensively on magneto-hydrodynamic stability theory(5). There were plans to move all the C.T.R. work to Winfrith, where ZETA II would be constructed. At A.W.R.E. both experimental and computational work began on the design of a uranium blanket to surround the plasma and so exploit the 14 MeV neutron flux(6).

This hybrid type of device could still prove to be effective.

It is worth remarking at this point that, although ZETA hit the headlines first in a very positive way and then rather negatively quite soon afterwards, its performance exceeded the design estimates and it produced a wealth of diagnostic information which is still being analysed. The experiment revealed a quite unexpectedly favourable phenomenon, namely the self-reversal of the B_ϕ field at the outside of the plasma followed by a quiescent period of enhanced stability. This is the basis of the Reversed Field Pinch (R.F.P.) containment line now being pursued at Culham, Los Alamos and Padua. Unfortunately, the phenomenon was not understood when ZETA was closed in 1968 and did not become clear until several years later when J.B. Taylor developed a minimum-energy principle in 1974(7).

Most fusion physicists at that time expected thermonuclear power to come into widespread commercial use much sooner than we do now (the present earliest estimate is about 2030), even though not all would have agreed with the over-optimism engendered by the early ZETA results. Military and civil technology had both advanced rapidly since the beginning of World War II. Sputnik had just been launched. Thus it was natural to expect C.T.R. to develop quickly also. In retrospect it seems clear that controlled fusion is unlike other technologies and that it justifies Walter Marshall's description as the most difficult technological project that man has attempted! Most other technical devices have had small beginnings - often small enough to be financed by a single 'back-room' inventor - and have then increased progressively in size as the success of each stage provided the financial justification for the next. By contrast, it appears that a toroidal magnetic containment system must be large in order to retain both the α-particles and the thermal energy sufficiently well to allow it to ignite and that as a result the minimum cost of a demonstration reactor is at least \$1000M.

At the beginning of 1959 I transferred to the Theoretical Physics Division at Harwell, then led by Dr. Lomer, to join Bill Thompson's plasma physics group, which was assisting the C.T.R. Division under Thoneman, Pease and Bickerton. There were two main interests in the group at that time, M.H.D. stability theory and transport theory. The very first C.T.R. concepts had neglected stability and, in the case of the stellarator and the toroidal Θ-pinch, even equilibrium; but it had soon become clear that in view of the microsecond M.H.D. timescales both were essential. The stellarator at Princeton was twisted into a figure-of-eight or given a rotational transform to ensure equilibrium, a toroidal stabilizing field was added to the Z-pinch at Harwell, while the toroidal Θ-pinch at A.W.R.E. was postponed for a time and replaced by a linear version, escape of plasma from the ends being accepted.

Most of the calculations on the toroidal pinch were made in cylindrical geometry which is a good approximation, so that it is still often referred to as a Z-pinch. The earliest stability calculations (Taylor[8], Rosenbluth[9], Shafranov[10]) employed a global surface-current model with uniform B_z and plasma pressure inside the surface and with vacuum B_Θ and B_z outside. These calculations predicted ZETA to be stable, which indeed it was for gross M.H.D. modes, but experiment also showed a high level of M.H.D. turbulence together with rapid cross-field energy and particle loss. By the time of the Second Geneva Conference a more sophisticated approach had been pursued by Suydam(11) and Rosenbluth(12) which was based on the ideal M.H.D. energy principle of Bernstein et al(13); a full theoretical prescription for the one-dimensional case was later given by Newcomb(14) and programmed by Copley and Whiteman(15). This allowed the pressure and the magnetic field to be continuous functions, p(r), $B_\theta(r)$, $B_z(r)$ and predicted additional localized instabilities. The subject of M.H.D. stability theory then became very complicated and remains so even now, partly because the eigenfunction and eigenvalue structure is singular and partly because the localized eigenmodes are influenced by non-ideal effects such as resistivity, viscosity and finite Larmor radius.

I was motivated at that time (and still am) by the need to analyse in detail the working of a plasma apparatus or reactor in detail as a function of time. It seemed reasonable to concentrate on the θ-pinch operated by Niblett's group at A.W.R.E. because of its simplicity, the single field B_z being a scalar in one-dimensional (r) and two-dimensional (θ,r) calculations. Bryan Taylor had embarked on a zero-dimensional theoretical analysis of this device, later with some help from me; it seemed a straightforward matter to optimize the implosion and subsequent heating phase and to achieve a high temperature - our slogan was 'a kilovolt by Christmas' (1958). The difficulty was that the initial conditions were not well understood, being determined by an uncontrolled preionization phase. Later, of course, much higher temperatures were reached.

To a mathematical physicist it seemed obvious that the initial conditions would be important and I remember discussing the early stages with Raymond Whipple and Ernest Laing. The experimentalists simply fired the bank without worrying how the ionization process would start; yet if one considered the short time involved there was unlikely (for example) to be even one cosmic-ray electron present. There was, however, a strong electric field at the start which might cause field-emission depending on the wall conditions. Occasionally the discharge did not fire at all on the first half cycle. This made prediction difficult.

A.W.R.E. experience suggested that a one-dimensional simulation would be appropriate, but Harwell had only a Mercury computer. However, during the summer of 1959 Klaus Hain from the Max Planck Institut für Physik und Astrophysik at Munich visited Harwell for an extended period. He had not only been involved with the M.H.D. stability theory of Hain, Lüst and Schlüter(16) but was also a computing expert, this expertise growing from the wartime German work of Zuse carried out in parallel with the U.S. and U.K. efforts. He had developed a one-dimensional M.H.D. code on the early computer at Munich and had presented a paper at Uppsala(17), but it was clear that the boundary conditions were not quite right. I was used by that time to implosion and explosion hydrodynamics and, although the boundary conditions for a plasma are obscure, I had read a paper by Colgate, Ferguson and Furth(18) which gave a prescription for the injection of pressureless plasma at the wall which would evidently do. Furthermore we had access to the fast and well-organized IBM709 and 704 computers at A.W.R.E. and Risley. Klaus and Gerti Hain, Sheila Roberts and I therefore embarked on a full scale one-dimensional M.H.D. code to model not only the fast linear θ-pinch but also fast stabilized Z-pinches such as Tarantula and slow stabilized Z-pinches such as ZETA. This was an exciting time - we were the first substantial users of Fortran outside the U.S.A. and at times had the IBM704 entirely to ourselves. The code(19) worked well and agreed with fast-pinch experiments and, although for minor numerical reasons it did not do all that had been planned, it was widely used for many years afterwards in various versions not only at Harwell, Culham and Munich but also elsewhere in Europe and in the U.S.A. Sheila Roberts, David Fisher and I improved the code organization while Joanna Taylor added a neutral gas component with ionization and charge-exchange effects. It taught us a great deal about computational physics, numerical analysis, code structure and documentation and the proper organization of scientific computing systems in general as well as about the behaviour of fast θ-and Z-pinches. These extensive plasma computations led to the formation of the Culham Computing Group and were used as part of the justification for the acquisition of the I.C.T. Atlas and the formation of the ATLAS Computing Laboratory, although, in the event, Culham and Harwell continued to use the IBM STRETCH at A.W.R.E.and subsequently acquired their own English Electric KDF9 and IBM360/65 computers respectively. The successful demonstration of a large Fortran program led to a rapid growth of interest in Fortran at Harwell, culminating in the implementation of a Fortran compiler for Atlas.

The transfer of C.T.R. work from Harwell and Aldermaston to Culham under its new Director, J.B. Adams, began in 1961. The theoretical plasma group at Harwell was amongst the first to move, partly in order to obtain its own computing facilities organized by Bill Morton and Leon Verra and partly to be near the first experiments. Culham had originally been planned with the pinch as its major line and the large D1 building was designed to hold the big Intermediate Current Stability Experiment (I.C.S.E.) which was a thin-skin pinch. I.C.S.E. was, however, abandoned in view of theoretical considerations which showed that such a configuration would not be stable, and of computer calculations which suggested that it was unlikely to be established in any case. In one respect this was fortunate because we were able to build a diversity of small experiments and to develop a wider range of theories. On the other hand if we had known enough to build a diffuse-current, reversed-field pinch as a successor to ZETA the pinch line might have been 20 years further ahead.

Until 1961 interest lay in the linear Θ-pinch which in its reversed field version was achieving kilovolt temperatures - although it was recognized that because of its open ends it could hardly form the basis of a thermonuclear reactor. This line was pursued at Los Alamos, N.R.L. in Washington, Culham and elsewhere; in fact Artsimovich remarked at the Salzburg conference that 'every housewife has her own Θ-pinch'. Toroidal pinches and stellarators appeared to show Bohm diffusion and mirror machines exhibited the predicted flute instabilities. Then at Salzburg a new era began with the announcement of an important result by Ioffe: the stabilization of the mirror by the use of additional conductors (Ioffe bars) which altered the symmetry and produced a true field minimum or 'magnetic well' at the centre of the plasma rather than a saddle-point as hitherto(20). First results from the Princeton Model C stellarator were also announced at Salzburg.

Salzburg started off a number of new lines. It had finally been shown that M.H.D. instability could be conquered - at least in the mirror machine - and groups at Livermore, Culham, Fontenay and elsewhere quickly fitted Ioffe bars or their equivalents. We did not understand the field structure at first; I remember that we produced some crude stereoscopic diagrams with an early graph plotter and studied them with a binocular viewer, while Dick Post brought some stereoscopic slides from Livermore which we examined through spectacles. Finally, Mike Larkin and I constructed a large cardboard model from computer-generated plots of the $|B|$ contours and field lines. (It was completed at home on the evening when President Kennedy was assassinated.) This model made everything clear. Curiously enough, a photograph of the model - which looked very elegant - was reproduced in a press advertisement by a company dealing in metals. Other better, wooden models were built in the workshop. More significantly, Taylor developed a general theory of magnetic well stabilization(21) and Larkin showed that instead of the mirror coils and Ioffe bars a single coil could be employed, shaped like the seam of a tennis ball (or, in the U.S.A., a baseball). Later two Yin-Yang coils were used. Princeton work on the stellarator led to similar work in the U.S.S.R., U.K., Germany and later Japan. Another theoretical line that emerged from Salzburg was finite-Larmor-radius stabilization, first proved by Rosenbluth, Kroll and Rostoker(22) by Vlasov methods but later shown by Taylor and myself(23) to be derivable from additional terms in the M.H.D. equations. The resistive instability theory of Furth, Killeen and Rosenbluth(24) was also developed at about this time. Taylor and I showed(25) that their highly localized g-modes could be replaced by equivalent, non-localized quasi-modes which might be much more dangerous in the non-linear regime - although the significance of this was not appreciated by anyone (including ourselves) for many years.

For the next few years fusion prospects gradually declined. Bohm diffusion seemed unavoidable in the stellarator and the Model C did not perform well, while turbulence and rapid loss of particles and energy occurred also in the pinch. It became clear by accurate computation that if longitudinal electric fields were allowed for the end loss in a mirror reactor by classical collisions would make the Lawson product too small; research on mirrors thus eventually ceased except at Livermore and in the U.S.S.R. Theoreticians discovered a host of new micro-instabilities in every type of confinement system, as well as enhanced 'neo-classical' diffusion in toroidal systems resulting from the finite size of the particle drift orbits in a non-uniform magnetic field.

For such reasons the fusion programme at Culham was reviewed and a decision was made to reduce it by half over a five-year period. Thoughts had in any case been turning to diversification from 1964 onwards, stimulated by Harold Wilson's "white heat of technology", and one of the successful early areas was computing. Culham and Winfrith had ordered KDF9 computers from English Electric and had jointly designed the EGDON operating system which was implemented by the company with remarkable speed and skill. Peter Poole then joined the Culham Computing Group under Bill Morton and rapidly developed the COTAN on-line terminal system which proved highly popular. The term 'software engineering' was coined at Culham at about that time. As a consequence of the Flowers Report(26), A.E.A. computing standards spread throughout the universities and the Culham hardware configuration together with the EGDON-COTAN operating system was adapted by all university and many government and industrial KDF9 installations. The GHOST graphical system was also introduced at Culham and widely marketed, while Niblett developed the STATUS information retrieval system and made valiant efforts to persuade the legal profession to use it.

Relief to the hard-pressed fusion programme came from Moscow. Good results had been reported for some time from Tokamaks, which are toroidal pinch-like devices with a high B_Φ then used only in the U.S.S.R., but they were not theoretically understood as the diagnostics were incomplete. As the result of a strong initiative by Bas Pease, who was now Director of Culham, it was agreed with the U.S.S.R. at the 1968 Novosibursk Conference to send a Culham diagnostic team (Peacock, Robinson, Forrest and Wilcock) and equipment to Moscow to measure the electron temperature and density on the Tokamak T3 by the new technique of laser Thomson scattering which had been developed on ZETA. I remember being in the U.S.A. while the measurements were in progress and being told that all contracts from Washington were held up until the results were known. When they proved(27) even more favourable than the U.S.S.R. had claimed and it became clear that an electron temperature of several hundred eV could be achieved in quite small Tokamak devices by ohmic heating, there began a massive switch to the Tokamak line which was led by Princeton, who quickly and successfully converted the Model C stellarator into the STC Tokamak.

The Tokamak line has since then become popular throughout the world, culminating in the four large devices, TFTR, JET, JT60 and T15, which are now under construction, and in the planning of INTOR under the auspices of the I.A.E.A. International collaboration has progressively increased since declassification in 1958; thus while JET was designed and is being built at

Culham by the nine Community nations together with Sweden and Switzerland, INTOR is currently being planned as a world project.

As the result of a great deal of theoretical and experimental work since 1969 the Tokamak is now much better understood. It depends for its MHD stability on being a relatively tight torus, unlike the pinch which can be an infinite cylinder. An important parameter is the safety factor or inverse rotational transform q(r); the number of turns of a field line the long way round the torus for each turn the short way. An MHD instability eigenmode must fit within the torus, i.e. in cylindrical approximation it must have the form $f(r) \exp(im\theta + in\Phi)$ where m and n are integers. Except near a 'singular surface' where $q(r) = n/m$ the field lines are distorted by the eigenmode and tend to stabilize it. Stability is favoured by a high shear (variation of q with r) and also by a mean magnetic well due to toroidicity.

The function q(r) normally increases with r as a result of the concentration of the current in the hot central region, but it must exceed unity everywhere if a (1,1) M.H.D. eigenmode is to be avoided. As the plasma heats up a thermal instability can occur in which the central temperature rises, q(o) falls below unity, and a redistribution of the inner region of the plasma takes place. This is seen as a sawtooth relaxation oscillation and is relatively benign, but a major disruption which appears to be caused by the overlap of two non-linear instabilities with different (m,n) can destroy the plasma completely and must be avoided.

The first Tokamaks used ohmic heating which can produce electron temperatures up to about 1 kev, but since the resistivity decreases as

$T_e^{-3/2}$ it becomes less effective at high temperatures and most workers agree that auxiliary heating will be needed in Tokamak reactors to achieve ignition. The injection of high-energy neutral beams was first employed in mirrors and during the planning of a Culham stellarator programme following the 1961 Salzburg meeting Pease suggested that it should also be used in toroidal devices. To some people at that time this seemed impracticable because of the low beam powers then achieved, but subsequently there has been a remarkable thousand-fold increase in power as the result of computational analysis, experimental diagnosis and improved engineering of the injectors. Many megawatts of neutral beam power are now available; these allowed the ion temperature to be raised to 5.5 keV on the Princeton PLT Tokamak in 1978. Radio-frequency heating seems to be following the same upward path.

When it became clear that, despite the broad magnetohydrodynamic stability of Tokamaks, their performance is determined by cross-field plasma and energy transport together with heating and radiation, one-dimensional evaluation codes were introduced(28) to study their behaviour as a function of time. Many physical effects have now been incorporated into these codes including neutral beam injection, impurity diffusion and the emission of neutral atoms from the wall. Comparison with experiment shows that the ion transport coefficients are approximately neo-classical but there is an anomalously large electron heat conduction which is not yet fully explained.

The economic performance of a Tokamak reactor cannot yet be guaranteed because the ratio β of plasma to magnetic pressure is only a few per cent, while the power output for fixed magnetic field is proportional to β^2.

Efforts are therefore made to calculate the maximum theoretical β for stability and equilibrium, to compare this with experiment, and to optimize β by shaping the torus cross-section and the plasma and current profiles. The currently-favoured cross-section is D-shaped (as in JET) and this requires two-dimensional equilibrium, stability and evolution codes which are now available. Instability modes with high (m,n) cannot, however, be treated by numerical simulation, and for these Taylor and co-workers have developed an elegant analytic method(29) which is related both to the W.K.B. approximation and to the quasi-modes mentioned earlier(25).

Fusion reactor studies were stimulated by Carruthers, Davenport and Mitchell(30) in 1967 and are now an active field with frequent conferences. The problems are immense and are not solely concerned with the plasma. One of the principal concerns is the first wall whose function is not to keep the plasma in but to keep the air out. This is bombarded not only by neutrons but by diffusing hot plasma, X-rays, γ-rays and fast charge-exchange neutral atoms. Particle bombardment can not only damage the wall but also knock off heavy atoms which penetrate the plasma and cool it by radiation. The economic feasibility of the reactor is dominated by the maximum energy flux (MW/m^2) at the first wall, while its practicability depends on being able to dismantle and re-assemble quickly this interior component even though it is surrounded by topologically-complex magnet coils and the intensely radioactive cooling circuit and blanket. The neutron flux per unit of power output is much greater than in fusion reactors. Furthermore, the neutrons have 14 MeV energy so that the induced radioactivity in the blanket is large, although there are no radioactive fission products and thus much less afterheat. Other problems under consideration include fuel injection and impurity and 'ash' removal, control of the ion temperature (since a reacting plasma is thermally unstable at its optimum operating temperature), the use of a divertor to protect the first wall, and thermal cycling of the blanket caused by a sequence of reaction pulses.

Other lines that have been studied at Culham and are still contenders for fusion power are the mirror, stellarator, pinch and inertial confinement. The mirror is a steady-state device which would have many engineering advantages if the problem of its low Lawson product could be solved. For this reason it is the main support line for the Tokamak in the U.S.A. and is being actively pursued at the Lawrence Radiation Laboratory at Livermore. One possibility is to put a uranium-lithium blanket around a mirror reactor turning it into a fission-fusion hybrid; another which is used in the so-called 'tandem' mirror, is to improve the longitudinal electrostatic potential distribution (which normally assists the escape of the ions) by placing auxiliary mirror machines at either end.

Early stellarator problems appear to have been solved by careful and accurate design of the magnetic fields using extensive trajectory calculations and by increasing the shear. Although the largest stellarator is much smaller than the largest Tokamak, a stellarator appears to behave at least as well as a Tokamak of the same size and possibly better. A stellarator reactor would have the advantage of operating in steady state although its lack of azimuthal symmetry makes both the theory and the engineering more complicated. There has been a revival of interest in stellarators since the Berchtesgaden Conference in 1976 although it is not yet clear whether CLEO (Culham, due to be discontinued in 1980) and Wendelstein VIIA (Garching) will have successors.

There has also been a revival of interest in slow reversed-field pinches of the ZETA type, considerably encouraged by Taylor's minimum-energy principle of 1974(7). Fascinating physics is involved in the fact that the plasma appears to seek and find its own stable minimum-energy state, physics which appears to be related to the solar and terrestrial dynamo problems. In addition to the old ZETA results there is evidence from fast pinches such as HBTX1 and FRSX at Culham for the correctness of the Taylor principle, but the magnetic Reynolds number is expected to be important so that fast and slow pinches should behave in qualitatively different ways. The new ETA-BETA II experiment at Padua has recently provided confirmation of the ZETA results and it is expected that ZT-40 at Los Alamos and HBTX1A at Culham will yield further information about slow pinches and the Taylor mechanism. A definitive test can, however, only be made with the RFX experiment proposed by Culham (radii 180/60 cm, currents up to 2MA) and it is hoped that this will go ahead as part of a Culham-Los Alamos-Padua collaboration. The RFP line may have reactor advantages over the Tokamak but this can only be established by an experimental test.

An area not taken up at Culham is laser fusion. The writer examined this possibility in 1964 but the field remained classified until 1972 when the results of computer calculations were published by Livermore. This was followed by theoretical and computational work at Culham culminating in the publication of the one-dimensional MEDUSA code(31) which has been extensively used in several countries. A proposal for a joint A.E.A.-S.R.C. laser fusion programme was, however, not accepted, and laser compression work is now carried out at the Rutherford Laboratory and at A.W.R.E.

What of the future? Because of their size, long construction timescale and cost (at least $1,000M for a device the size of INTOR) it will be essential that future large-scale devices work as planned. This involves a great deal of theoretical calculation beforehand, requiring in practice the use of elaborate computer codes whose parameters are carefully normalized to fit available experimental data on the behaviour of existing devices and which are then used to design and predict the performance of the next generation. It must, however, be emphasized that although the fundamental equations governing the behaviour of the plasma are completely known (Newton, Maxwell, Schrödinger) the number of degrees of freedom is so large that these equations can never, even in principle, be solved by purely numerical methods; the number of particles in the universe would be quite insufficient to build the computer required. A combination of simplified theoretical models, experimental measurements and computer calculations is required. This is somewhat different from the situation pertaining in the design of fission reactors or weapons and is more akin to that in hydrodynamics where one cannot hope to follow the turbulent flow of a fluid at high Reynolds number by purely numerical means.

The need for large-scale computation has led to the establishment of the National Magnetic Fusion Energy Computer Centre (N.M.F.E.C.C.) at Livermore in California, linked by high-speed data lines to the major U.S. national, industrial and university fusion research centres and by telephone, ARPANET and TYMNET to all other U.S. fusion research teams. N.M.F.E.C.C. is equipped with CRAY-1 and CDC-7600 computers and enables theoretical and computational fusion physicists and engineers throughout the U.S.A. to take part in a continuous collaborative programme, building and sharing a common library of codes. Another computer centre has been established at Nagoya in Japan and a CRAY-1 computer has recently been installed at Garching in Germany. Sponsored by the I.A.E.A., discussions have taken place on the formation and publication of an international library of fusion codes.

A critical factor is that small theoretical groups such as those at Culham and JET cannot any longer hope by themselves to develop and maintain all the codes that will be needed for the analysis of complex devices such as JET and INTOR. Furthermore, it is impracticable to obtain such codes from abroad on magnetic tape since they take time and effort to install, particularly on a different type of computer, and then rapidly become out of date. Our policy must therefore be to encourage the formation of a European Fusion Computer Network, similar to that of N.M.F.E.C.C., and at the same time to establish efficient data links both with the U.S.A. and also with Japan. In accordance with this policy it has been possible for some time to contact the major U.S. fusion centres from Culham terminals via ARPANET, and this U.S. link will shortly be strengthened and supported by other connections to Garching and with the plasma physicists working in the U.K. universities.

The future for theoretical fusion physics is therefore expected to be an active and continuous collaboration between teams from several countries working together on international projects such as JET, TFX and INTOR and supported by powerful computers and data links.

1. J.D. Lawson, Proc. Phys. Soc. 70, 6 (1957).

2. Nature, 181, 217-233 (1958).

3. W.B. Thompson, 'An Introduction to Plasma Physics', (Pergamon Press, Oxford, 1962).

4. W. Marshall, AERE Reports T/R 2247 (1957), 2352 (1958), 2419 (1960).

5. R.J. Taylor, AERE Lectures L102-6 (1959).

6. J.W. Weale, H. Goodfellow, M.H. Taggart and M.L. Mullender, Reactor Sci. Technol. 14, 91 (1961).

7. J.B. Taylor, Proc. Fifth I.A.E.A. Conf. on Plasma Physics and Controlled Nuclear Fusion Research (1974) paper IAEA-CN-33.

8. R.J. Taylor, Proc. Phys. Soc. B70, 1049 (1957).

9. M.N. Rosenbluth, U.S.A.E.C. Report LA-2030 (1956).

10. V.D. Shafranov, J. Nucl. Eng. 5, 86 (1957).

11. B. Suydam, 2nd Geneva Conf. 31, P/354 (1958).

12. M.N. Rosenbluth, 2nd Geneva Conf. 31, P/347 (1958).

13. I.B. Bernstein, E.A. Frieman, M.D. Kruskal and R.M. Kulsrud, Proc. Roy. Soc. 244, 17 (1958).

14. W.A. Newcomb, Ann. Phys. 10, 232 (1960).

15. D.M. Copley and K.J. Whiteman, Plasma Physics 4, 103 (1962).

16. K. Hain, R. Lüst and A. Schlüter, Z. Naturforsch, 12a, 833 (1957).

17. K. Hain, Proc. Fourth Intern. Conf. on Ionization Phenomena in Gases, IVA, (Uppsala 1959), p.843.

18. S.A. Colgate, J.P. Ferguson and H.P. Furth, UCRL Report 5086.

19. K. Hain, G. Hain, K.V. Roberts, S.J. Roberts and W. Koppendörfer, Naturforsch, 15a, 1039 (1960).

20. Yu. B. Gott, M.S. Ioffe and V.G. Telkovsky, Nucl. Fus. Suppl. Part III, 1045 (1962).

21. J.B. Taylor, Phys. Fluids 6, 1059 (1963).

22. M.N. Rosenbluth, N.A. Krall and N. Rostoker, Nucl. Fus. Suppl. Part III, 143 (1962).

23. K.V. Roberts and J.B. Taylor, Phys. Rev. Lett. 8, 197 (1962).

24. H.P. Furth, J. Killeen and M.N. Rosenbluth, Phys. Fluids 6, 459 (1963).

25. K.V. Roberts and J.B. Taylor, Phys. Fluids 8, 315 (1965).

26. A Report of a Joint Working Group on Computers for Research, H.M.S.O. Cmnd. 2883 (1966).

27. N.J. Peacock, D.C. Robinson, M.J. Forrest, P.D. Wilcock and V.V. Sannikov, Nature 244, 488 (1969).

28. Y.N. Dnestrovskii, D.P. Kostamorov and N.L. Pavlova, Proc. Int. Symp. on Closed Confinement Systems (1969); C. Mercier and Soubbaramayer, Ibid. (1969). For other references up to 1976 see: J.T. Hogan, Math. Comp. Phys. 16, 131 (1976); M.L. Watkins, M.H. Hughes, K.V. Roberts, P.M. Keeping and J. Killeen, Ibid. 165 (1976).

29. J.W. Connor, R.J. Hastie and J.B. Taylor, Phys. Rev. Lett. 40, 396 (1978).

30. R. Carruthers, P.A. Davenport and J.T.D. Mitchell, Culham Report R-85 (1967).

31. J. P. Christiansen, D.E.T.F. Ashby and K.V. Roberts, Comput. Phys. Commun., 17, 271 (1974).

Chapter IX

NEUTRON TRANSPORT THEORY AND OTHER APPLICATIONS

J.H. Tait

1. Introduction

This section deals with some of the work of the groups which J.H. Tait has been associated with from 1954 to the present. In chronological order these groups are the Neutron Transport Theory Group, the Atomic Physics Group, which later broadened its interests and became the Atomic and Molecular Physics Group, and finally the Nuclear, Atomic and Molecular Physics Group, after a further amalgamation with A.M. Lane's group. The work on atomic physics is described separately in articles by P.G. Burke and J.S. Briggs, except that part of this article (by David Hodgkinson) is on laser isotope separation - one of the current activities of the group - which introduces other aspects of atomic and molecular theory. Developments in Nuclear Reaction Theory are described by Tony Lane in Chapter IV.

The following contains an historical account of the work from 1954, starting with reactor physics theory and finishing with that relating to isotope separation studies.

2. Reactor Physics Studies

The Neutron Transport Theory Group was part of the Division in 1954 and existed until 1959 when, following the establishment of A.E.E. Winfrith, it was disbanded. J.H. Tait then spent a year in Reactor Division managing a technical office for the water-moderated reactor project. He returned to start the Atomic Physics Group in 1960. K.T. Spinney and R. Royston joined the Reactor Division, A. Hassitt transferred to Risley and P. Schofield joined W.M. Lomer's group.

The work of this group evolved from the activities of the Division prior to 1954. The main study in 1946 was Pile Theory, and a group under C.A. Rennie carried out the reactor physics calculation for the original Windscale reactors. This work was transferred to Reactor Physics Division on its formation and Theoretical Physics Division then concentrated on longer term work which was directed by B. Davison. The transfer of the project work to Reactor Physics Division coincided with a broadening of interests in T.P. Division and in 1954 its activities covered many fields besides that of reactor physics.

Although the group was called the Neutron Transport Theory Group, its activities covered a study of a wide range of reactor physics topics. These included studies of (i) neutron shielding (K.T. Spinney[1]), (ii) fast reactor multi-group calculations (Betty Mandl) - which were the first to be carried out in the Authority, (iii) the development of two-dimensional codes for the solution of the neutron diffusion equation (A. Hassitt), (iv) the design of the booster target for the electron linear accelerator (R. Royston), (vi) neutron thermalisation theory (P. Schofield) described in Chapter X and (vi) fast reactor safety studies (J.H. Tait). R.T. Whipple was also a member of the group during part of this time and worked on the calculation of energy yields in criticality accidents in chemical plant. J.H. Sykes also worked for some of the time on neutron transport theory.

One piece of work will be highlighted, namely that relating to the study of the whole core accidents in fast reactors. It is chosen because it was studied over many years in the Division and finally there was a fruitful U.K.-U.S. collaboration culminating in the work of H.A. Bethe and J.H. Tait. The problem had been considered prior to 1954 by B. Davison, M.H.L. Pryce, R.T. Whipple and J.H. Tait and rather large estimates were obtained for the energy release in a fast reactor meltdown accident. This problem was reviewed in discussions with a U.S. fast-reactor delegation, of which H.A. Bethe was a member, when it visited Harwell. A new calculation was made and the results reported to a conference on reactor safety held in Chicago in 1956(2). This calculation initiated a major study, mainly in this country and the U.S., of the energy yield following the loss of coolant and meltdown of a fast reactor core.

In order to obtain an overestimate of the energy release it was assumed that all the coolant was lost from the core of the reactor, which became molten instantaneously and settled freely under gravity. The reactivity change as the core settled freely under gravity was calculated using perturbation theory, and the rate of reactivity addition at prompt critical determined. Again in order to obtain an overestimate, it was assumed that the molten core behaved as a superheated liquid and that no outward expansion occurred until the voids left by the coolant had been filled by the internal expansion of the superheated liquid. This meant that a certain amount of energy per unit volume Q^x had to be released before any outward expansion occurred. The theory resembled that of a detonation wave in a chemical explosive and it was possible to calculate the energy release analytically.

Many extra effects have since been added to the calculation, e.g. the Doppler effect (first reported in fast reactors in 1958), the sodium void coefficient and fuel-coolant interactions with the result that the models are now more realistic. However, the first calculation brought the problem to the attention of the reactor designers, and the original estimates are correct to an order of magnitude.

Although the Neutron Transport Group was disbanded in 1959, it was started up again in a modified form under P. Schofield. J.H. Tait continued an interest in reactor problems while managing the Atomic Physics Group. Work was carried out on the design of a super-booster target for the electron linear accelerator. Other ways of producing intense fluxes of thermal neutrons were investigated, e.g. by the use of a high energy accelerator. This followed some earlier work carried out in collaboration with K.T. Spinney in the 1950's. A considerable amount of effort was devoted to the study of the diffusion of species produced in the radiolysis of carbon dioxide in the pores of the graphite moderator of an A.G.R. reactor. In the 70's a joint working group was set up with Culham to study safety problems in fast reactors, e.g. fuel-coolant interactions and core-catcher design - the Division provided both the Chairman and Secretary.

Over many years the Division has provided a service to the Harwell Criticality Committee. J.H. Tait and K.T. Spinney in the 50's carried out many safety calculations and the Division has provided the Chairman for the past eighteen years.

3. Isotope Separation Studies

Theoretical Physics Division's interests in isotope separation started in 1948. The Establishment was looking into the possibility of using the high speed gas centrifuge for the enrichment of uranium but little was known concerning the flow pattern in such machines. Two excellent mathematicians, B. Davison and R.T. Whipple, tackled this problem. Davison analysed the flow in terms of eigen-functions which exist in a very long cylinder. Whipple, however, produced a solution in terms of boundary layer flows and this was particularly relevant to machines working at higher pressures. This solution was very elegant and clarified the physics of the problem. Whipple later extended this work and produced a report in 1962, which is now declassified[4]. It is unfortunate that the declassification was delayed until 1978 because similar solutions have been published elsewhere and Whipple has not had full credit for his work internationally. The solution of the flow pattern in a centrifuge is a very complicated one and at the time (1948) it was interesting to watch the competition between Davison and Whipple. Although Whipple won on this occasion, Davison's approach provided the germ of an approach adopted later by J. Hubbard and J. Tait. The interests of the Authority are described in Heinz London's book on the separation of isotopes[3].

After this early excursion into the field of isotope separation interest waned, and it was not until 1968 that work was started on the flow of gas through diffusion-plant membranes. This was carried out by J. Tait and J. Beeby and later Joanna Taylor helped.

The U.K. adopted centrifugation as the technique for uranium enrichment in the early 70's, and the Division was asked to carry out calculations of the flow and separation in modern high speed machines. This work was started by J. Tait, who was later joined by J. Hubbard and Joanna Taylor. A semianalytical method was used incorporating a modification of the Whipple theory for flow over the end caps, but in the main part of the machine the flow was described in terms of eigen-functions for a very long cylinder, the so-called Steinbeck functions. Coupled with the separation calculation the method is fast and extremely useful for survey calculations of the output as a function of various machine parameters. The Division is still active in this field through the efforts of Mair Williams and recently P. Stopford has joined the team. In 1970 the Treaty of Almelo was signed and the theoretical work in Germany, Holland and the U.K. was co-ordinated by CENTEC. There were enjoyable meetings at Bensberg with our European colleagues. With the amalgamation of CENTEC and URENCO the direction of some of our work has come directly from B.N.F.L.

In 1974 work on laser isotope separation was started in the Division by J. Tait. J. Briggs helped in this work and we were later joined by D. Hodgkinson, P.T. Greenland and Joanna Taylor. Although the effort is small, contributions have been made to the project. Some of the work is unclassified and in a later section some of the contributiuons made by Briggs and Hodgkinson are described.

Work has also been carried out on the assessment of the potential of other techniques and, for example, the foam process was analysed with the help of J. Beeby, now a consultant. Studies have also been made of the separation of stable isotopes of interest to R.C.C., Amersham.

4. Laser Isotope Separation (D.P. Hodgkinson)

During the last decade, new methods of isotope separation based on laser excitation of atoms and molecules have been proposed and developed. The idea is that the laser's narrow line width can be exploited to excite a single species of an isotopic mixture while its high power allows a large fraction of the selected species to undergo a permanent physical or chemical change before dissipative and scrambling processes occur. For uranium the promise is that energy costs could be reduced a hundredfold and capital costs tenfold in comparison with other technologies[5] and that uranium reserves could be extended through a near complete separation of isotopes.

The traditional rate theory of absorption and stimulated emission has no validity when high power lasers are used as the radiation source. Instead it is necessary to treat the interaction to all orders in the electric field of the laser. This leads to phenomena such as multiphoton transitions, power broadening of absorption lines and the associated Rabi oscillations of excited state occupation probabilities.

4.1 Atomic route

The most straightforward realisation of these ideas is the selective excitation and ionisation of an atomic vapour by two or more lasers. One of the problems with this method is that the power-broadened line width is generally a small fraction of the Doppler width and only this small fraction of atoms is therefore excited. A way around this difficulty is to use two oppositely directed laser beams such that their Doppler shifts cancel out, so producing an overall two-photon resonance for all thermal velocities. The consequences of this suggestion for the strong field conditions required in laser isotope separation were worked out by Hodgkinson and Briggs[6] and this early work has subsequently been extended to incorporate many of the complexities found in uranium spectra. Once the desired isotope has been successfully photo-ionised it can undergo a resonant charge exchange collision with the unexcited species thereby losing the hard-won isotopic selectivity. This problem stimulated an extensive theoretical study of resonant charge exchange cross-sections by the above authors(7).

4.2 Multiphoton dissociation of molecules

One of the most exciting developments in the field of laser isotope separation was the discovery of isotopically selective multiphoton dissociation of polyatomic molecules in an intense infra-red laser beam(8). In this process, molecules with a vibrational absorption band close to the laser frequency rapidly absorb in excess of thirty quanta and dissociate without the aid of collisions. Moreover, it is isotopically selective as witnessed by one of the early experiments(9) where enrichment factors of 2,800 were observed. Isotopes of a large number of elements have been enriched in this way on a laboratory scale and recently this has been scaled up to the grams-per-day level for sulphur hexafluoride.

The interesting theoretical question in collisionless multiphoton dissociation is how so many photons of equal energy are rapidly absorbed despite the fact that the anharmonicity of the molecular vibrations makes transitions between high-lying levels in the infra-red active mode non-resonant. Certainly power broadening(10,11), anharmonic splitting(12) and rotational sub-structure(13) of the vibrational levels can partially compensate for the anharmonic defect. However, at the reported intensities, and using realistic molecular parameters, it is possible to explain the absorpition of only a few quanta(14) rather than the thirty or more required for dissociation.

A way out of this impasse is to include the effects of intramolecular coupling betweeen the normal modes due to anharmonic terms in the potential energy. The harmonic modes are dynamically independent at small displacements but when the mode coupled to the laser becomes significantly excited energy can leak from it into the rest of the molecule by this mechanism. When this happens the excitation of the absorbing mode decreases and so it does not reach a high level of excitation thus circumventing the problem of the anharmonic defect.

A theoretical framework based on these physical ideas was developed by Hodgkinson and Briggs(15) and simultaneously by Cantrell at Los Alamos(16). They established a formalism for the calculation of the absorption of energy from an intense electric field by a collection of interacting anharmonic quantum oscillators, one of which is coupled to the applied field. This is analagous to problems encountered in quantum statistical mechanics where it is found convenient to describe such a system by a generalised master equation for the reduced density matrix of the absorbing mode.

Approximate solutions to this equation were examined for a simplified set of vibrational levels and it was demonstrated that the above explanation is plausible for relatively small coupling widths(11). At the time there was no reliable way to estimate these, but recently spectroscopic evidence(17) has been interpreted(18) as confirming their approximate magnitude.

Recent work in the field has concentrated on improving the model of the molecular states, for example by including the effects of rotation, anharmonic splitting and vibration-rotation interaction[19]. With these refinements the theory predicts dissociation at lower intensities than first envisaged, as found experimentally.

1. B.T. Price, C.C. Horton and K.T. Spinney, Radiation Shielding (Pergamon Press, 1957).

2. H.A. Bethe and J.H. Tait, An Estimate of the Order of Magnitude of the Explosion when the Core of a Fast Reactor Collapses. Paper for Reactor Hazards Meeting, Chicago, 1956. RHM(56)/113.

3. Separation of Isotopes, Ed. H. London (George Newnes Ltd., 1961).

4. R.T.P. Whipple, Private Communication (1962).

5. J.H. Birely, D.C. Cartwright and J.G. Marinuzzi, Proceedings of the SPIE Seminar in Depth on Ultra-High-Power Lasers for Practical Applications, 1976.

6. D.P. Hodgkinson and J.S. Briggs, Opt. Commun. 22, 45 (1977).

7. D.P. Hodgkinson and J.S. Briggs, J. Phys. B: Atom. Molec. Phys., 9, 255 (1976).

8. R.V. Ambartzumian, V.S. Letokhov, E.A. Ryabov and N.V. Chekalin, JETP Lett., 20, 273 (1974).

9. R.V. Ambartzumian, Yu. A. Gorokhov, V.S. Letokhov and G.N. Makarov, JETP Lett., 21, 171 (1975).

10. D.M. Larsen and N. Bloembergen, Opt. Commun. 17, 254 (1976).

11. D.P. Hodgkinson and J.S. Briggs, Chem. Phys. Letts., 43, 451 (1976).

12. C.D. Cantrell and H.W. Galbraith, Opt. Commun. 21, 374 (1977).

13. R.V. Ambartzumian, Yu. A. Gorokhov, V.S. Lesokhov, G.N. Makarov and A.A. Puretzkii, JETP Lett. 23, 22 (1976).

14. J.R. Ackerhalt and H.W. Galbraith, J. Chem. Phys. 69, 1200 (1978).

15. D.P. Hodgkinson and J.S. Briggs, J. Phys. B: Atom. Molec. Phys., 10, 2583 (1977).

16. C.D. Cantrell, Laser Handbook, Vol. III (North-Holland Publishing, 1977).

17. A.S. Pine and A.G. Robiette, J. Mol. Spectrosco., to be published, 1980.

18. J.R. Ackerhalt and H.W. Galbraith in Laser Spectroscopy, IV, Eds. H. Walther and K.W. Rothe, (Springer-Verlag, New York; Heidelberg, Berlin) to appear, 1979.

Chapter X

NEUTRON PHYSICS AND THE THEORY OF LIQUIDS 1956-1976

P. Schofield

I joined Harwell in October, 1956 as a member of John Tait's Neutron Transport Theory Group. This period coincided with the assessment by the A.E.A. of the design tenders for the first Magnox power reactors and my first task was part of this. It was to calculate the effects upon the reactivity of the reactor core brought about by changes in the energy spectrum of the thermal and epi-thermal neutrons resulting from changes in fuel composition and temperature within the core. The early graphite-moderated reactors at Hanford, Windscale and elsewhere had been designed for the production of plutonium and this plutonium was removed from the core before it had a significant effect on reactivity. For the purposes of calculating reaction cross-sections, the thermal neutron spectrum could then be characterized by two parameters - a Maxwellian temperature and an amplitude of the '1/E' tail. By contrast, the new designs of reactor involved going to much higher burn-up so that the spectrum was affected by the plutonium fission resonance at 0.3 eV. In addition, one design proposed a sleeve of graphite to support the fuel; this would necessarily be at a much higher temperature than would the bulk moderator and would thus 'heat' the neutrons entering the fuel.

Similar problems were arising in the design of both water-moderated and heavy-water-moderated reactors, so there was at that time a growing world-wide activity on the rather new subject of 'neutron thermalisation'. The theoretical problem divided into two distinct parts. One was the determination of the inelastic scattering cross-sections of moderating materials for thermal neutrons. The second was the solution of the neutron transport equation, given these cross-sections. On the experimental side, there were three principal developments. Peter Egelstaff set up the 'Scattering Law Project' on the NRX reactor at Chalk River in 1957 to measure inelastic scattering cross-sections directly. Michael Poole in Hanger 8 and eventually also on the LINAC booster - for which design calculations were carried out by Royston in Theoretical Physics Division - measured thermal neutron spectra in moderating material by using time-of-flight techniques. Graham Campbell, in Reactor Physics Division at Harwell and later at Winfrith, used foils to obtain integral measurements over the spectrum in various assemblies.

The first spectrum calculations at Harwell were carried out by Tony Hassitt and myself on graphite containing uranium and plutonium fuel; these were presented at the second Geneva conference on the Peaceful Uses of Atomic Energy in 1958(1). The main problem was the calculation of the inelastic cross-sections for graphite - at low energy the one-phonon term dominates, but at epi-thermal energies multi-phonon terms dominate, so that a 'short-collision time' type of approximation is appropriate. The two extremes were reconciled by applying the central limit theorem to the phonon expansion of the cross-section and by showing that, at high energy, this reduced to a steepest-descent evaluation of the integral which occurs in the formulation of the scattering law.

Extension of this theory to liquid moderators was the next step. Vineyard, at the Brookhaven National Laboratory, had, in 1958, introduced the Gaussian approximation to the incoherent scattering law. Egelstaff and I then showed how, within this approximation, the scattering depended on the spectral function of the velocity auto-correlation of the scattering atoms, which in the solid corresponds to the phonon density of states. Given the spectral function (which could be obtained from the Chalk River data), it was therefore possible to use the same methods to calculate the cross-sections. These were, in fact, incorporated in two computer codes, LEAP and PIXSE, by Bob McLatchie. These are still in use today and are currently being used to study the possible performance of metal hydride moderators for pulsed sources of neutrons, in particular the new Harwell LINAC and the planned Spallation Neutron Source at the Rutherford Laboratory.

In 1962, the Neutron Physics Group was formed (later to become the Neutron and Liquid Physics Group). Originally it consisted of myself, David Wilmore, who was making optical-model calculations of neutron-nuclear cross-sections for the reactor programme, and Phil Hutchinson, whose main interest was the statistical theory of fluids, but who contributed also to the scattering law work. My own preoccupations remained with neutron thermalization and the scattering law, but this gradually tailed off as the calculations became more routine. The methods devised then now form part of the major reactor physics design codes used by the Authority. The last major work on thermalization was with Mike Lancefield, a graduate student from St. Andrews University, on the thermal neutron Milne problem, i.e. the spectrum of neutrons emerging from a plane surface(2).

In parallel with neutron thermalization, an interest was developing in the use of thermal neutron scattering as a probe of the structure and dynamics of condensed matter. The commissioning of DIDO and PLUTO in the late fifties provided neutron beams of sufficient intensity to obtain good quantitative information on the structure (i.e. the pair distribution function) and atomic motion in liquids. In particular, with Egelstaff, we studied models for the diffusive motion of atoms in liquids, as measured by quasi-elastic neutron scattering.

In 1963, Walter Marshall and I considered (in parallel with other groups throughout the world, notably Mori in Japan) the scattering from hydrodynamic fluctuations in liquids, thus giving a microscopic basis to the old Landau-Placzek theory(3). The analogous theory for light scattering from plasma led to the introduction of optical methods as major diagnostic tools in fusion work.

Temporary members of the Group in the 1960's worked on various aspects of radiation scattering from condensed matter and related problems. These included P. Michael (Brookhaven) on neutron scattering from methane (including proton spin correlation), M. Tanaka (Tokyo) on general scattering problems, T. Högberg (Studsvik) on anharmonic phonon theory and J. Ranninger on heat transport in crystals. A most significant visitor was George Benedek from M.I.T., who spent six months here in 1966 as Harwell's first Professorial Fellow. Benedek had pioneered the use of laser light scattering to study the diffusion of macromolecules in solution and with Elizabeth Bradford we worked out the theory of scattering for anisotropic molecules with coupled rotational and translational diffusion(4). But the major outcome of Benedek's visit was the initiation of the use of laser Doppler techniques to study turbulent fluid flow.

I remember well the lunch-time discussion with Egelstaff, Benedek and Bill Denton and Pat Bourke of Chemical Engineering Division at which this possibility was first mooted. The first experimental measurements of the turbulent intensity were carried out, in collaboration with R.R.E., Malvern, and published in 1968[5]. Early exploitation of the technique at Harwell was delayed through a decision to concentrate effort on commercial instrument development. However, the importance of the technique to the Heat Transfer and Fluid Flow service, and particularly to the study of flow in furnaces (where high temperatures rule out the use of conventional methods) led to a programme of work in collaboration with Imperial College, under Phil Hutchinson's supervision. Since then, many applications have been found for laser Doppler shift measurements. These are now concentrated in Les Drain's section in Materials Physics Division. They include an important component of the Internal Combustion Engine project and also, at last, a rig for studying turbulence in simple geometries to provide test data for the finite-element flow modelling (see chapter XI by John Rae).

Following Benedek's visit to Harwell, I was invited to M.I.T. in 1968, and while there was able to devote some time to the theory of critical phenomena, which until then had remained a peripheral interest. Inspired by the beautiful experiments of Ho and Litster[6] on the transparent ferromagnet chromium tribromide, I proposed[7] a parametric form for the equation of state close to a critical point. This was based on Widom's homogeneity hypothesis for the dependence of the free energy on 'scaled' variables. It turned out that the equation of state could be represented by a nearly (though not exactly) linear dependence on one of the parameters for a number of magnetic and fluid systems. I understand that this equation of state is now recommended in the chemical engineering literature for estimating the critical behaviour of 'real' fluids! A few years later, the theory of critical phenomena was transformed by K.G. Wilson's introduction of the renormalisation group(8). In 1972, John Hubbard and I were the first to apply this to the liquid-vapour critical point(9) and, in the process, we discovered, albeit unknowingly, what later became known as 'redundant variables'!

During the 1960's, as in many other fields, computer simulation methods began to play an increasingly important rôle in the study of the liquid state, mainly through the pioneer work of Alder and of Rahman. In 1968, Dave Beeman, a temporary research associate from California, wrote the Harwell molecular dynamics programme. This was further developed by Norris Dalton and myself. The aims of the work were to explore the relationship of the interaction potential to the structure of and atomic motion in simple fluids, both in order to test the theory and to aid the interpretation of neutron scattering experiments. The initial studies were on the relative rôles of the short-range repulsion and of the long-range attraction between atoms in determining liquid state properties. In fact, the generated data proved extremely valuable in validating the perturbation methods of calculation (based on a purely repulsive 'reference system') which were becoming popular at that time(10).

As an alternative to the perturbation method, Hutchinson and Conkie (who came to us on attachment from Queen's University, Ontario) developed a thermodynamically consistent theory for equation-of-state calculations. In an extended version, developed in collaboration with Brennan and

Sangster (from Reading University), this provided a means of obtaining interatomic potentials from liquid state diffraction data(11). Earlier attempts by Hutchinson with Johnson and March (then at Sheffield University) to do this by using the non-consistent Percus-Yevick and Hyper-netted chain approximations had failed to give the 'hard-core' radius correctly. Swapan Mitra (on attachment, but supported by contracts with Imperial College and later Oxford University) took on the more difficult problem of estimating effective interionic potentials in metals from the liquid state data(12). This work continues to-day, with continually improving theoretical understanding. Such progress would not have been possible without the ability to test the results against the exact calculations provided by the computer simulation method.

The Harwell molecular dynamics program was the basis for further developments by various other Groups. Sangster and Dixon at Reading extended it to deal with molten salts(13) and their program, in turn, became the basis for the important work of Gillan (who joined Theoretical Physics Division in 1970) and Dixon on clarifying the mechanism of ionic conduction in the fluorite lattice(14). At Studsvik, Sweden, I collaborated with Ebbsjö, Waller and Sköld to study isotope effects in a model of Argon(15). This work led further, in collaboration with Turq and Lantelme (Université de Paris VI) to the study of the isotope effect on the mobility of ions in molten salts(16). It is clear that the molecular dynamics simulation method has an assured future in unravelling increasingly complex phenomena in solids and liquids.

To end with a personal comment, I would say that the rôle of Theoretical Physics Division to-day is much closer to that when I joined it than it has been in some of the intervening years, particularly in respect of the strength of its links within Harwell and elsewhere in the A.E.A. However, this is to a large extent a consequence of long-term research initiated during the 1960's coming to fruition. The importance of the Division lies in its ability to attract the best theoreticians by providing an atmosphere where new ideas can be discussed and developed against a clear view of long-term objectives, but without an overriding emphasis on the short-term.

1. P. Schofield, and G. Hassitt, Proc. 2nd Geneva Conf. on Peaceful Uses of Atomic Energy (1958), p.18.

2. M.J. Lancefield and P. Schofield, Brit. J. Appl. Phys. 18, 1497 (1967), J. Phys. D. 1, 137 (1968).

3. P. Schofield in "Physics of Simple Liquids" (North-Holland, Amsterdam, 1966), Ch. 13.

4. D.W. Schaeffer, G.B. Benedek, E. Bradford and P. Schofield, J. Chem. Phys. 55, 3887 (1971).

5. P.J. Bourke et al, J. Phys. A 3, 216 (1970).

6. J.T. Ho and J.D. Litster, Phys. Rev. Letts. 22, 603 (1969).

7. P. Schofield, Phys. Rev. Letts. 22, 606 (1969).

8. K.G. Wilson and M.E. Fisher, Phys. Rev. Letts. 28, 240 (1972).

9. J. Hubbard and P. Schofield, Phys. Letts. 40A, 245 (1972).

10. J.D. Weeks, D. Chandler and H.C. Anderson, J. Chem. Phys. 54, 5237 (1971).

11. M. Brennan, P. Hutchinson, M.J.L. Sangster and P. Schofield, J. Phys. C 7, 2411 (1974).

12. S.K. Mitra, P. Hutchinson and P. Schofield, Phil. Mag. 34, 1087 (1976).

13. M.J. Sangster and M. Dixon, Adv. Phys. 25, 247 (1976).

14. M. Dixon and M.J. Gillan, J. Phys. C 11, L165 (1978), AERE reports TP.794 and TP.809 (1979).

15. I. Ebbsjö, P. Schofield, K. Sköld and I. Waller, J. Phys. C 7, 3891 (1974).

16. F. Soutelme, P. Turq and P. Schofield, Mol. Phys. 31, 1085 (1976), J. Chem. Phys. 67, 3869 (1977), J. Chem. Phys. 71, 2507 (1979).

Chapter XI

FLUID BEHAVIOUR

John Rae

1. Introduction

In Theoretical Physics Division (T.P.D.) a strong interest in fluid mechanics began only quite recently. It roughly spans the period of the diversification policy at Harwell and, as will be seen later, is closely associated with non-nuclear projects. Current research reflects this diversification and embraces turbulent flow, combustion, oil reservoir modelling and nuclear waste migration as well as the invention and development of computer methods.

The starting point lies in the late 1960's, when Phil Hutchinson* began studies for the Heat Transfer and Fluid Flow Service (H.T.F.S.) of the motion of liquid drops in two-phase flow and, at about the same time, there was a growth of interest in the theory of turbulence, coinciding with the successful development of a laser anemometer which could measure instantaneous fluid velocities. Both these lines of work have continued to the present day and have an influence on current work, as mentioned below. The third main topic from fluid mechanics arose in the early 1970's with John Truelove's work on radiative heat transfer in gas furnaces. This aspect also survives strongly today although only a part falls under this division. Subsequent developments, particularly those related to the finite element method, began in 1974 and led directly to the varied fluid mechanics work of current projects.

* Now in Engineering Sciences Division.

2. Annular Two-Phase Flow

The justification for research into two-phase flow is quite straightforward. This type of flow occurs in a vast range of engineering equipment (e.g. evaporators, condensers, etc.); it is a very complex physical phenomenon and is very poorly understood. The commonest regime of two-phase flow, known as annular, is typified by a central stream of gas in the middle of a duct with a moving film of liquid - annular for a pipe - on its walls[1]. The exchange of droplets between the gas stream and the film has a great effect on heat transfer. Early on it was realised that droplet production was closely allied to the motion of large amplitude waves on the liquid film, and after Hutchinson's move to Thermodynamics Division (as it then was) it was these waves which were studied in T.P.D. The amplitudes of the waves are very large, several times the film thickness, and they can only be modelled by non-linear theories involving difficult modern mathematics.

3. Turbulence

The problems of turbulence have been known for a long time. The complexities of turbulent motion were, in fact, illustrated by Leonardo da Vinci[2] and, despite formidable theoretical attacks in the last hundred years by Taylor[3], Kolmogorov[4] and others, a resurrected Osborne Reynolds[5] would notice only modest progress. One way of seeing the difficulties is to consider the scales of space and time which arise in turbulent flow. Turbulence is commonly pictured as a cascade process in which a shear flow or stirring mechanism pumps energy into large eddies. The energy then spills through smaller structures down to scales where it is dissipated quickly by viscosity. This picture and some dimensional analysis can show that the smallest length scales, 1, and shortest times, τ, are related to the lengths, L, and times, T, of the driving shear flow through the Reynolds number, R_e, by

$$1 \simeq \frac{L}{R_e^{\frac{3}{4}}} \text{ and } \tau \simeq \frac{T}{R_e^{\frac{1}{2}}} .$$

In real flows the Reynolds number is typically 10,000 to 100,000 so in turbulence one has to accommodate a continuous range of lengths and times varying by perhaps three orders of magnitude. The lack of a clear separation of scales rules out perturbation methods of the sort which are so successful in kinetic theories of fluids, although many of the commonly used turbulence models, in fact, follow that line. The range of scales also precludes the direct calculation of turbulence even with the best modern computers. Thus, for a three-dimensional calculation one would need computational grids of perhaps 10^{12} blocks and a correspondingly huge number of timesteps.

In the face of these difficulties turbulence theories have therefore evolved into two distinct lines. There are mathematically "rigorous" theories which avoid ad hoc assumptions but are restricted to ideal forms of turbulence, and there are rough empirical mathematical models which are applied, with limited success, to real engineering flows. The greater part of the work in Theoretical Physics Division has attempted to evaluate some models from this second line and the computer programs in which they are used. Difficult computational questions arise, however; for example, problems of 'numerical diffusion' which has a spreading effect similar to turbulence. These show the need for carefully developed and evaluated methods.

4. Radiative Heat Transfer

A related set of problems arose from John Truelove's work on radiative heat transfer. In order to calculate the radiation in a gas furnace we need to know the temperature distribution and hence the flow pattern of the turbulent gas. Initially, vain attempts were made to buy in a ready-made program for this but later, through the efforts of John Sykes and others, it proved possible to build on an Imperial College program and eventually, in 1979, to issue the extensively developed and modified program TUFC through H.T.F.S. When Truelove, and later Sykes, moved to Engineering Sciences Division in 1975 and 1976 the mainstream program development went with them. However, after the oil crisis of 1973 and setting up of the International Energy Agency in 1974, oil conservation became more important and T.P.D. began work on combustion in oil-fired furnaces. Mathematical models have been constructed for the motion of oil droplets in a furnace and for the processes of evaporation and combustion. These processes and the flow of air and waste gases couple together very strongly and provide a difficult problem of computation. The current calculations use an extended form of TUFC which solves fifteen coupled non-linear partial differential equations(6). We are also aware of the potential application of these methods to other technical areas, such as engine development, and tentative first steps have been taken towards these applications.

5. Applications of Finite Element Methods

In 1974 computational fluids work in T.P.D. took a totally new direction with our interest in the finite element method of solving differential and partial differential equations. Our expertise in this is one of the characteristics of the Theory of Fluids Group, which is one of only five or six such groups in this country. The finite element method evolved originally as a numerical technique for stress analysis and structural calculations (including heat conduction) during the 1960's and in that context its mathematical validity depended on the formulation of these problems as variational principles. Many important equations of physics, for example the Navier-Stokes equations for viscous flow, cannot be formulated in this way and it was not until the early 1970's that it was realised there were other techniques, such as the Galerkin formulation, which permitted the method to be applied to other field problems. Since then its use has spread rapidly[7,8]. Enthusiasts for the method claim that it gives higher accuracy and a natural treatment of curved boundaries and local mesh refinements. Traditionalists using finite differences have claimed that these advantages are illusory and have been spurred into improvements of their own methods. A healthy debate continues and - however it may end - it is clear that computational physics will come out richer.

In this Division the development of the finite element method for flow began on orthodox lines with studies of steady laminar flows in reasonably complicated geometries[9]. Since then we have built carefully, adding the capabilities to perform time-dependent studies, heat and mass transfer and chemical and other rate equations. We are now at the point of returning to turbulent flow studies in which we can use our carefully evaluated programs to discriminate between effects coming from numerical inaccuracies associated with the method of discretization and true features of the turbulence model itself. Most members of the Theory of Fluids Group have contributed to developing these programs, known collectively as WIFE, and the work continues at a modest level of effort. In the meantime, however, new needs had arisen and our computational fluid mechanics had branched off in another and unexpected direction.

With the discovery of North Sea oil the government and industry became increasingly interested in oil and gas production methods and in building up relevant expertise. The widely held view of oil being produced in the form of controlled gushers is very misleading; it is, in fact, difficult to extract oil and a thirty per cent success rate is considered good. Sophisticated production strategies are made for well positions and rates of extraction, also for the position and rates of wells for water injection or gas re-injection to maintain pressure. As geological and physical data are never plentiful, especially for under-sea fields, various approaches are used in working out the strategy. One such is the technique of mathematical modelling and, in fact, the application of this technique in the oil industry has been responsible for many of the advances in computational fluid mechanics in the last twenty years(10). In the circumstances it is not surprising that Harwell became involved in work of this type, first through the Department of Energy and later, additionally, with the British Gas Corporation and British National Oil Corporation. At first the work lay mainly in the Computer Optimisation Group (now in C.S.S.D. but part of T.P.D. until 1973) and involved the development of a finite-difference simulation program. But the oil and gas industries had at that time just begun to investigate finite-element techniques, so the Theory of Fluids Group was ideally placed to help as specialists in this method.

We are now engaged in writing a finite-element simulation program, called RESOLVE, for two fluid phases (oil and water or gas and water) in a two-dimensional reservoir. We are also conducting related research into advanced techniques. The underlying physics is that of two-phase flow in a porous rock medium. Although there exist dispersive processes in this system the fronts or interfaces between the phases are often very sharp in relation to the size of the reservoir and require to be tracked very carefully. Naïve methods usually give wrong front speeds and again involve numerical diffusion, with its consequent artificial spreading of the front.

The benefits of designing software flexibly and with built-in potential were shown recently with the rise in interest in underground nuclear waste disposal. One of the problems in this area is coupled heat and mass transport by underground water flow, driven both by natural hydraulic gradients and by convection from the heating. Superficially there is some similarity to oil reservoir problems but the basic physical mechanisms are quite different. In less than six months in 1979 we put together on a basis of WIFE routines a new and quite sophisticated finite-element program for combined heat and groundwater flow. It is now being used to study processes of importance in waste repositories(11). Our background of experience in flow and mass transfer calculations has been of great help in designing radio nuclide migration models to supplement the flow equations and is another of our current activities.

6. Conclusion

In summing up, one may say that work in this Division on large scale fluid behaviour has grown from very small beginnings about ten years ago to occupy the full-time efforts of some half a dozen scientists today. In the main our calculational methods are built round the WIFE package of finite element programs which we develop and maintain to a standard as high as any in the world, but this does not preclude use of other methods when appropriate. Our real strength in this work, however, lies in the ability to grasp the necessary physics for setting up models and to analyse their results in a physically useful way. This is exemplified in the applications to gas and oil-drop combustion, turbulent flow, oil reservoir simulation and nuclear waste repositories and will, undoubtedly, carry into new fields as other problems arise.

1. G.F. Hewitt and N.S. Hall-Taylor, Annular Two-phase Flow (Pergamon Press, Oxford, 1970).

2. Drawings of Leonardo da Vinci, Collection of the Institut de France Bibliotheque, Paris.

3. G.I. Taylor, The Statistical Theory of Turbulence, Parts I-IV, Proc. Roy. Soc, A151, 421 (1935).

4. A.N. Kolmogorov, The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds Number, C.R. Acad. Sci. U.R.S.S., 30, 151 (1941).

5. O. Reynolds, An Experimental Investigation of the Circumstances which Determine Whether the Motion of Water shall be Direct or Sinuous, and the Law of Resistance in Parallel Channels, Trans. Roy. Soc. (London) A174, 935 (1883).

6. K.A. Cliffe, D.A. Lever and K.H. Winters, A Finite Difference Calculation of Spray Combustion in Turbulent, Swirling Flow, AERE-R9507 (1979).

7. D.H. Norrie and G. De Vries, The Finite Element Method (Academic Press, New York, 1973).

8. J.J. Connor and C.A. Brebbia, Finite Element Techniques for Fluid Flow (Newnes-Butterworth, London, 1976).

9. K.A. Cliffe, C.P. Jackson, J. Rae and K.H. Winters, Finite Element Flow Modelling Using Velocity and Pressure Variables, AERE-R9202 (1978).

10. H.B. Crichlow, Modern Reservoir Engineering, a Simulation Approach (Prentice-Hall, Englewood Cliffs, 1977).

11. J. Rae and P.C. Robinson, NAMMU-Finite Element Program for Coupled Heat and Groundwater Flow Problems, AERE-R9610 (1979).

Chapter XII

ELECTRON BAND THEORY 1952-1962

W. M. Lomer

1. Introduction

I started at Harwell in 1952, coming straight from Cambridge, recruited specifically to join Metallurgy Division to form a small theoretical group. When I reached the Pass Office, I was told that 'they' had changed their minds and wanted me to join T.P. instead. T.P. at that time was apparently headless, heving just 'lost' Fuchs, and Brian Flowers was not yet in post. Senior consultancy was provided one day a week by Maurice Pryce from Oxford and by Rudolph Peierls from Birmingham. They, of course, were not there on the relevant day, but the Division Head's Office existed, and Megan Kenyon at least knew which desk I was to occupy, and where the pencils and paper were. Then, in a short but stilted 'chat' Boris Davison, the senior man present, asked me "Do you have everything?" and "Do you have a problem to work on?" With both questions answered affirmatively there was nothing more to say, and nothing to do but start. In the course of the next week, I had met Derek Johnston, the only other inhabitant of T.P. who was being encouraged to follow up the theory of solids and, of course, had re-established contact with my many acquaintances in Metallurgy Division.

The problem Johnston was engaged on was the interpretation of measurements of the electrical properties of radiation-damaged graphite; the problem I wanted to start on was the detailed core structure of dislocations in real materials following up the semi-quantitative work I had done with Bragg on the 'bubble model' of crystals[1,2]. It seemed obvious from that work that the core structure of a dislocation was calculable if one could determine a suitable two-body interaction and carry out a relaxation calculation to find the equilibrium positions of the atoms.

Within a few months two more recruits had been found for the solid state section, Alan Foreman and Roy Leigh. Foreman brought a formidable mathematical capacity in elasticity theory and Leigh a thorough knowledge of electron band theory learned at the feet of Harry Jones. The strategy for 'my' problem was simple now. We would all concentrate on copper as a model material. Leigh and I could work out the energy as a function of lattice constant by the methods of band theory, and match them to suitable Morse potential curves, while Foreman could match up the elastic dislocation field to some local relaxation calculation round the core region where elasticity theory locally failed. The strategy for graphite was less obvious, although it was clear that the theory of electrical conduction of semi-metals needed detailed calculation of the band structures, but methods were not well established. Two things emerged rather quickly; firstly, that a two-dimensional single sheet of graphite - a hexagonal network of carbon atoms - ought to have an electronic structure similar to the real thing and, secondly, that the theory of the representation of space group states would simplify the calculations both for the real substance and for the single sheet.

These were the elements of the programme at that time, and it is interesting to try to elucidate where the trail led, and how far along the ever-lengthening road my successors have taken matters.

2. Graphite

The method of the Linear Combination of Atomic Orbitals (L.C.A.O.) was by the mid-1950's a well established technique for describing the valence bonding schemes in molecules, and it gave good insight into the structure of the lower energy levels. Coulson and Taylor in 1952(3) had given an L.C.A.O. description of the conduction band of single sheet graphite, based on combinations of the p_z electrons on each atom. (Because these are anti-symmetric with respect to reflection in the plane, they are separate from the s, p_x and p_y states that form the bonding band.) In 1955(4) I wrote a lengthy paper on the L.C.A.O. basis for the valence band structure of two-dimensional graphite, in which the space group categorization of states was fully utilised to establish the sequence of states at high-symmetry points in the Brillouin zone, and to establish the connectivity of the bands along the symmetry lines. It was pointed out that the whole treatment was approximate and was more likely to be acceptable for the low-lying valence band states than for higher states, and that the production of numerical estimates for the several parameters rested on a series of semi-empirical fits to observation and to other calculations in a quite alarming way. One has to remind onself, looking back from this computer-rich age, that the computations of a two-centre overlap integral between even simplified nodeless functions was a considerable undertaking; as a result we all borrowed liberally from any appropriate sources to avoid that hard labour. In the same way, reduction of a 3 x 3 matrix eigenvalue problem to a 2 x 2 and a single equation was a major step, saving as it did the heavy computational cost of solving cubic equations!

The conclusions of the paper that were noted were directed at the light which the two-dimensional calculation shed on the three-dimensional band structure. The main result was that parts of the bonding valence states lay within one volt of the Fermi energy, so that mixing of the conduction band states based on p_z electronic states with valence states might be important for the three-dimensional solid. A side comment was made about the fact that Wannier functions formed by summing over individual bands were not like the Pauling sp^2 hybrids, even though the same principle of maximum overlap was used to form each. The fact that space group theory for a lattice with a basis had yielded results with a minimum of computation was not specifically mentioned, though I believe this was the first paper to use the technique in an L.C.A.O. calculation.

At the same time, Derek Johnston used the L.C.A.O. model to evaluate the energy of a band based exclusively on p_z orbitals to discuss the possible shape of the Fermi surface in graphite(5). There are four atoms per unit cell in graphite, so that all the secular equations were of dimension four. This calculation gave Fermi surfaces that were cylindrical around the edges of the Brillouin zone, and of very small dimensions indeed. At all practical temperatures the majority of carriers - holes in the valence band and electrons in the very slightly overlapping conduction band - were the result of thermal excitation. The effects of radiation on the measured Hall coefficient were interpreted in terms of the trapping of electrons at defects, so that the Fermi level fell and the holes increasingly dominated the transport properties. On the whole, the model was a good one, but the temperature variation of the Hall coefficient was not well described, so we attempted to provide a better description of the Fermi surface by taking account of π-σ mixing by enlarging the basis of the L.C.A.O. calculation to include all four atomic states on all four atoms in the unit cell(6). By classifying the states by their symmetry properties and studying their connectivity, we showed that the Fermi surface would consist essentially of a long thin cigar of holes, touching end-on a long thin cigar of electrons, these being located at the corners of the hexagonal Brillouin zone. We also toyed with the inclusion of spin-orbit coupling with the aim of finding details of the structure of the cigars which would be permitted by group theory but not exhibited within the limited L.C.A.O. basis we had used previously. This led ultimately to the publication of Johnston's magnum opus on Group Theory of Electrons in Crystals(7). But for graphite we found that the cigar structures became quite complicated when these extra parameters were included; for the symmetry operators now permitted energy levels which were degenerate at the zone corner to cross at a finite slope. The three-fold symmetry around the axis at the zone corner leads to elegantly fluted energy surfaces which are locally not analytical. Many years later the de Haas van Alphen technique and the general paraphernalia of Fermiology established many of these features, but at that time we could not put this detail into the transport quantities and the work just stopped.

Graphite, of course, remained a principal interest; the "Wigner energy", the stored radiation damage of clustered vacancies, interstitials or whatever, was first of academic interest and then, after the Windscale reactor fire, of acute practical concern. Estimates of vacancy formation energy, diffusion rates, defect densities and annealing mechanisms were all built up with experimental and theoretical contributions from John Simmonds, Dennis Rimmer, Alan Cottrell at Harwell, Charles Coulson and Simon Altmann in Oxford, and the group at R.D.L., Risley, as well as the T.P. group. Nowadays, graphite is mainly of interest for the chemical problems of its interaction with coolants and for the rôle of surface contamination and catalysis in that interaction. Indeed it was ultimately determined that this, even more than the Wigner energy, was responsible for the fire at Windscale.

3. Dislocation Core Structure: Electronic and Interatomic Forces

In 1950, "it was well known that" the elastic constants of copper could be estimated by methods used by Fuchs in which only three terms really counted - the free-electron-gas energy of the electrons, the Madelung energy of the singly charged ions in the uniform electron gas, and the closed d-shell interactions. Unfortunately, this gave a very poor value for the cohesive energy; because the closed-shell interaction was purely repulsive, copper was no more strongly bound than sodium, and the elastic constant was reasonable only because the interatomic spacing was imposed rather than derived as the most stable spacing. Furthermore, my own experience with calculating the elastic constants of the bubble rafts had shown me that Poisson's ratio could only be calculated if one knew about three-body forces. I puzzled for a very long time about how to calculate distortions of the closed shells under elastic stresses. P.-O. Löwdin and his collaborators in Sweden seemed to be on the right track with their calculations on alkali halides, but we never got far for copper. The required concepts of a d-shell polarizability and of multipole shell-model distortional parameters were all there waiting for enunciation, but we got nowhere trying to deduce it all from closed-shell overlaps, exchange repulsions, etc.

Meantime, however, Leigh was learning and teaching us about more detailed ways of dealing with the conduction electron part of the problem, working up the A.P.W. method and exploring its accuracy and convergence. A particular point of interest for copper was the possibility of a reduction of certain elastic shear constants which could occur when the energy of one pocket of electron states is raised and that of a symmetry-related pocket lowered by the elastic distortion. In this way the transfer of electrons from one set of states to the other would reduce the elastic constant. This turned out to be quantitatively uninteresting for copper, and interest in the matter lapsed. Much later this question has again been taken up when considering structural stability and soft phonons, etc. We were therefore left high and dry for the detailed dislocation calculation because of lack of understanding of the "closed shell" interactions and hence of all possibility of producing a sensible two-body force cystem. The problems of dislocation energy and slip plane choice were instead tackled by different semi-empirical methods(8).

The problem of assigning effective ranges to the interatomic forces continued to interest us, and in 1957 the interpretation of thermal diffuse scattering of X-rays and neutrons caught our attention. We showed(9) that in principle the range of the forces could be estimated from the phonon spectra along lines of high symmetry. This showed that ranges out to fourth or fifth neighbours had to be considered in aluminium, and left the problem of matching observations to a priori calculations still as far from solution as ever.

4. Uranium

Its nuclear properties aside, uranium is still funny stuff. Thus it has three structural phases:

(i) alpha - a base-centred orthorhombic structure that is, in fact, a slightly distorted close-packed hexagonal,

(ii) beta - a complex tetragonal phase (with 20 atoms per unit cell) rather similar to the well known intermetallic sigma phase, characteristic of intermetallic compounds, and

(iii) gamma - a body-centred cubic structure.

The physical properties of the alpha phase are highly anisotropic, and it seemed sensible to try to apply the methods of group theory to elucidate the band structure. The space group of the alpha-structure is very simple so that it does not provide many distinct representations. The electronic states were expected to be combinations of 7s, 7p, 6d and 5f atomic functions. It was quickly found that the group structure(10) did not help much. Likewise there existed at that time neither reasonable atomic functions to use in an L.C.A.O. calculation nor a self-consistent field to calculate them from. At about this time Cicely Ridley came from Douglas Hartree's group in Cambridge to work for us, and one of her early heroic tasks was to get the Ferranti Mark 1 computer at Aldermaston to give a Hartree (exchange-free) self-consistent field for the U^{6+} ion(11). That calculation involved many all-night runs with the machine being 'hand-fed' with the results from previous phases of the computation since all the memory space was always required for the phase in hand.

The U^{6+} core was to be the starting point for obtaining the atomic valence electron functions to use in an L.C.A.O. calculation. It quickly became apparent that the radial dependence of the valence atomic orbital depended drastically on the occupancy of the available states. Furthermore, spectroscopic evidence showed that the spin and orbital states of the whole atom controlled the configuration energy through exchange interactions in a way completely neglected in the Hartree method of calculation. Equally, it was obvious that although one could easily see how to work in the Hartree-Fock approximation, the computers available were quite inadequate to make any progress. Desperate remedies were needed. In metals, in general, Hartree fields seemed to work because the orbital quantization was suppressed by crystal fields, and also because in the non-magnetic metals the Bloch states were each occupied by two electrons with paired spins. We therefore tried calculating some Hartree self-consistent fields for neutral uranium with 'metallic' boundary condsitions(12). These boundary conditions, following the Wigner-Seitz tradition, were to make the radial gradient of the k = 0 wave-function zero at the surface of a sphere whose volume equalled the atomic volume of the crystal lattice. The functions were made self-consistent with a configuration $(5f)^2$ $(6d)^2$ $(7s)^2$, while the band width of the s, d, and f bands was estimated by also finding the energy (in the same potential field) which made the wave-function itself zero on the surface of the Wigner-Seitz sphere. It was postulated that this energy would be about three times further above the zero-gradient bottom-of-the-band value than the real band top would be, since the real band would be better approximated by a zero-gradient boundary condition in two dimensions and zero value in one dimension. (A numerical check on the d- and s-band widths in copper by this method and comparison with A.P.W. calculations showed the factor to be indeed very close to three for each band.)

By studying other configurations, i.e. $(5f)^6$, $(5f)^4$ $(6d)^2$, and $(6d)^6$ we showed that self- and mutual-screening effects could indeed lead to a stable predicted configuration, which turned out to be close to $(5f)^4$ $(6d)^2$.

Table 1

Energy Parameters for Metallic U

Configuration $(5f)^2$ $(6d)^2$ $(7s)^2$

	Wigner Seitz boundary condition	P = 0 boundary condition	Band Width
7s	- 0.35 a.u.	- 3.22 a.u.	12.9 eV
6d	0.175 a.u.	- 1.061 a.u.	5.58 eV
5f	0.393 a.u.	0.267 a.u.	0.57 eV

Configuration $(5f)^4$ $(6d)^2$

	Wigner Seitz boundary condition	P = 0 boundary condition	Band Width
6d	0.0859 a.u.	- 1.247 a.u.	6.01 eV
5f	0.0786 a.u.	- 0.121 a.u.	0.9 eV

This crude computatonal scheme based on exchange-free non-relativistic methods was forced on us by lack of computational power. Nevertheless, the work was instructive and left us convinced of several important conditions, namely:

(i) the 5f and 6d electrons were both equally active in determining the bonding and crystal structure of uranium,

(ii) the 7s and 7p states might be little occupied, the conduction electrons were mostly in 6d states,

(iii) self-consistent calculations should lead to an understanding of the relative band occupancy in metals,

(iv) Wigner-Seitz boundary conditions could shed much light on general band parameters,

(v) pre-formed bonding hybrid orbitals of the kind used by theoretical chemists in discussion of the shapes of organic molecules were not likely to fit into or be derived from any reasonable theory of metals.

Over the next few years Hartree-Fock and 'unrestricted Hartree-Fock' calculations for atoms reached a good standard of accuracy. As a result it became clear thet the width of the 5f-band gradually contracted between thorium and plutonium and that, whilst our detailed numerical estimates were not of much direct use, most features of our general picture remained valid(13,14).

5. Correlation, Exchange and Magnetism

The problem of the cohesive energy of metals is dominated by the need to allow properly for exchange and correlation between the mobile metallic electrons. Roy Leigh followed up the original ideas of Wigner with some nice work on variational estimates of the total energy in real band systems. Then in 1951 Bohm and Pines published their paper showing a way of decoupling the collective plasma oscillations from the individual particle degrees of freedom. John Hubbard joined in at about this point, with his paper relating the diagrammatic analysis of the many-body perturbation series to the dielectric theory of plasma oscillations. His later papers used the results to calculate the correlation energy of the free electrton gas, later extending it to cover the electron gas in a crystal. But John Hubbard is writing one of these articles, and all I need say is that his work put Harwell on the world map of many-body theory and kept it there for two decades.

One conclusion from this work, however, was that there was good reason to treat the Hartree field as basic, and the exchange and correlation energies together as a next stage of approximation. This gave us the feeling that one-electron band theory should indeed give approximate treatments of many metals. The lack of computing power at Harwell meant that from the late

1950's onward we watched whilst the American teams - especially at M.I.T. under Slater - developed and compared A.P.W., O.P.W., and other schemes of band calculations.

Harwell's attention became directed more particularly to the problems of magnetism, i.e. to the states of metals and insulators with net atomic spins and possibly net orbital moments. The growing power of neutron diffraction made this especially rewarding, and the work on magnetic materials by Walter Marshall, Alan Runciman, John Gabriel and Dennis Rimmer elucidated many of the niceties of the effects of crystal fields on atomic states. John Hubbard produced his famous treatment of the narrow band correlation problem which remains the basis of most calculations of the stability of magnetic metallic states.

6. Transition Metals

It was only in 1959 that John Wood[15] from M.I.T. published a rather complete A.P.W. calculation for paramagnetic metallic iron in both face-centred cubic and body-centred cubic phases. The b.c.c. calculation was of particular interest for several reasons:

(i) it was possible to deduce the energy splitting between the two spin directions (which was not unreasonably high) and to show from the density-of-states curve that the spin moment might well stabilise itself at about two Bohr magnetons

(ii) it was possible to assign symmetry labels to all of the electron states and establish that some of the bands were essentially d-like and that some were hybridised s-p states

(iii) it was possible to see that the Fermi surface was placed to give roughly one occupied s-state per atom.

In 1961 Bacon[16] had established that Cr was antiferromagnetic with a periodicity that was not quite commensurate with the crystal lattice. This problem fascinated me, since it implied a self-consistent field with two incommensurate periodicities, implying that it was formally impossible to describe the wave functions as Bloch functions at all. Overhauser had just published his papers[17] on spin-density waves in an electron gas, which suggested that a spatially periodic spin density of wave-vector corresponding to the diameter of the Fermi sphere, 2 k_F, might be self-stabilising, since the corresponding potential could couple states of similar energy whose wave-vectors differed by $2k_F$ and so lower the energy of the favourable linear combination linearly with the magnitude of the potential. More careful criticism of Overhauser's original suggestion showed that the free-electron gas was not unstable in this way. It seemed reasonable, nevertheless, to speculate that with a non-spherical Fermi surface things might be different and in particular that paramagnetic chromium might have some special instability that favoured its unique anti-ferromagnetism.

In 1962[18] I therefore stuck my neck out by constructing a Fermi surface for paramagnetic chromium by the simple expedient of using Wood's iron calculation as a basis, lowering the s states a little relative to the d-band, which I assumed to be rigid, and then allocating 6 electrons per atom to the band system. To my joy, it was immediately clear that large areas of Fermi surface were nearly parallel and separated by wave-vectors that could easily be supposed to match the observed periodicity of the antiferromagnetic lattice. The instability proved difficult to establish quantitatively, and my efforts in this direction ran into the sands, as recorded in the report of the Varenna summer school on magnetism[19]. The less complex calculation of the static magnetic susceptibility of chromium, in terms of induced orbital moments arising from interband magnetic coupling was, however, carried out by John Denbigh[20] on a simple desk-calculator during a two month summer visit, using those same band energies from Wood's calculation for iron. The result fitted experimental analyses very well, and lived for several months before a real calculation with a computer superseded it. Later, computer calculations of the wave-vector dependent susceptibility became practicable, and somewhat to everyone's embarrassment, showed that the simple heuristic arguments were misleading; the susceptibility does not show nearly so sharp a peak as expected.

7. Conclusion

The Theoretical Physics Division between 1952 and 1965 devoted considerable energy to understanding electrons in metals, with uranium and the actinides and the structurally-important transition metals as the main targets. Many importrant insights were obtained. A main result of that period was a conviction that the majority of the physical properties of all metals, except the 4f rare-earth series and the actinides beyond uranium, were dominated by band effects which could be described well enough for most purposes by simple one-electron calculations with simple self-consistent fields. The period from 1960 on showed increasingly clearly the necessity of incorporating relativistic spin-orbit coupling terms in the heavy metals, and some 'local exchange field' correction to the fields close to nuclei. Recent work[21], however, shows that Th is still a tetravalent 6d metal, exactly as we deduced from Cicely Ridley's non-relativistic work on U in 1956! The problems of the non-local interaction of spins - highly important for alloy theory and for antiferromagnetic instability - required the evolution of computers large enough to produce wave-functions at all wave-vectors for all bands so that the susceptibility at arbitrary wave-vector could be computed. This work has not proved to be very illuminating so far, and much interest again focusses today on heuristic arguments that give qualitative descriptions of band structures, such as canonical d-bands to account for crystal structure (Pettifor[22]). The struggle is still as much for insight and understanding as for quantitative explanation that rests on complex and impenetrable computation.

1. W.M. Lomer, Proc. Camb. Phil. Soc. 45, 660 (1949).
2. W.M. Lomer, Proc. Roy. Soc. A196, 171 (1949).
3. C.A. Coulson and R. Taylor, Proc. Phys. Soc. A65, 815 (1952).
4. W.M. Lomer, Proc. Roy. Soc. A227, 330 (1955).
5. D.F. Johnston, Proc. Roy. Soc A227, 349 (1955).
6. D.F. Johnston, Proc. Roy. Soc. A237, 48 (1956).
7. D.F. Johnston, Rep. Prog. Phys. 23, 66 (1960).
8. A.J.E. Foreman and W.M. Lomer, Phil. Mag. 46, 73 (1955).
9. A.J.E. Foreman and W.M. Lomer, Proc. Phys. Soc. B70, 1143 (1957).
10. D.F. Johnston, AERE T/M 132 and 133 (1956).
11. E.C. Ridley, Proc. Roy. Soc. A243, 422 (1957).
12. E.C. Ridley, Proc. Roy. Soc. A247, 199 (1958).
13. R.G. Boyd, A.C. Larson and J.T. Waber, Phys. Rev. 129, 1629 (1963).
14. T.J. Watson-Yang, A.J. Freeman, D.D. Koelling, Bull. Am. Phys. Soc. 23, 275 (1978).
15. J.H. Wood, Phys. Rev. 126, 517 (1962).

16. G.E. Bacon, Acta. Cryst. 14, 823 (1961).

17. A.W. Overhauser, Phys. Rev. 128, 1437 (1962).

18. W.M. Lomer, Proc. Phys. Soc. 80, 489 (1962) and 84, 327 (1964).

19. W.M. Lomer, Proc. International Summer School of Physics, Theory of Magnetism in Transition Metals (Academic Press, New York, 1966).

20. J.G. Denbigh and W.M. Lomer, Proc. Phys. Soc. 82, 156 (1963).

21. H.L. Skiver, to be published.

22. D.G. Pettifor, J. Phys. C3, 367 (1970).

Chapter XIII

MANY-BODY THEORY

J. Hubbard

When the Atomic Energy Authority was founded 25 years ago the discipline of many-body theory was in its infancy. The intervening years have seen it pass through childhood and into adolescence, but we do not detect any real maturity as yet. Much has been learnt, some difficult phases have been passed through and forgotten, but no unified identity has yet been established.

In the years preceding 1950 some notable progress had been made in the theory of interacting particle systems. In particular one might mention the highly developed perturbation theory of celestial mechanics, the classical kinetic theory of gases, the exact solutions to certain one-dimensional problems obtained by Bethe, and the 'mean-field' theories such as Hartree-Fock. However, problems such as the prediction of the detailed correlations brought about by the Coulomb interactions of electrons in metals had been widely regarded as intractable. But the early fifties saw the beginnings of some progress on these 'intractable' problems. Particularly noteworthy were Bohm and Pines' theory of plasma oscillations in electron gases(1) and Breuckner's theory of the interactions of nucleons in nuclear matter(2).

Bohm and Pines showed that the Coulomb interactions gave rise to high energy plasma oscillations in the electron gases of metals and that these plasma oscillations to a large extent 'used up' the effects of the Coulomb interactions, leaving the electrons otherwise free to behave as though they did not interact, i.e. according to the prescriptions of band theory. Breuckner showed that the 'hard-core' interactions of nucleons could be dealt with by t-matrix summations leading to a tractable theory resembling Hartree-Fock.

These developments sowed the seed for what was later to emerge as the discipline of 'many-body' theory. But at that time the necessary mathematical apparatus for dealing with such problems had not yet been found, and these theories were formulated in terms of rather special and unsatisfactory mathematical schemes. Thus at the time of the formation of the Authority in 1954 the search was on for a sound approach to such problems. The first success in this direction came in 1957 with the work of Bethe and Goldstone(3) in Cambridge, who, borrowing from quantum electrodynamics the Feynman diagram technique for the analysis of perturbation series, formulated the Breuckner theory in a satisfactory way and proved the first important general theorem of many-body theory, the 'linked-cluster theorem'.

In the meantime, progress had been made with the electron gas problem. Mott(4), Fröhlich(5) and Hubbard(6) pointed out in the mid-fifties the close connection between the plasma oscillations and the dielectric properties of an electron gas; this 'dielectric' viewpoint has since that time remained the preferred way of looking at the plasma oscillation theory. Hubbard, by then working in the Theoretical Physics Division, sought a proper mathematical formulation of this dielectric scheme, and eventually (in 1957) developed a scheme(7) similar to the Bethe-Goldstone theory, but now adapted to the electron gas problem. At about the same time a similar scheme was developed by Breuckner and Gell-Mann (stiff competition indeed), but their theory did not go quite as far as Hubbard's. This perturbation theory remains to this day one of the standard approaches.

Another major and quite independent development occurred at about this time. Since its discovery in 1911 the phenomenon of superconductivity had been a great unsolved mystery. It had been realised after the development of quantum mechanics that it must be a quantum phenomenon, and it was appreciated that electron interactions must be responsible (for what else could it be due to?). The observation of the isotope effect suggested that the interactions in question were not the direct Coulomb interactions between electrons, but a secondary interaction between them mediated by the phonons of the ion lattice. Theory predicted that the latter interactions could give rise to an attraction between electrons, and Cooper in 1956 showed that they might lead to a bound state between a pair of electrons. However, the effects of the Pauli principle interfered with the bound state in a subtle way. In 1957 Bardeen, Cooper and Schrieffer unravelled this difficult problem, and showed that the attraction could lead to a condensed state of the electron gas which would indeed exhibit superconducting properties(8). The mystery was solved! Again rather special techniques were employed in the initial formulation of the theory, but it was soon reformulated in terms of, and absorbed into, the main stream of many-body theory being developed at that time. The Bardeen-Cooper-Schrieffer theory has stood the test of time and remains one of the great successes of many-body theory.

The next few years (the late fifties and early sixties) saw a general flowering of the theory without startling new developments. The perturbation theory technique was augmented by the development of other techniques such as the Green's function methods of Zubarev and Martin and Schwinger, which have proved to be invaluable mathematical tools, but did not of themselves provide new physical insights. Various other techniques such as the equation of motion method and the functional integral method (Stratonovitch(9), Hubbard(10)) were also developed about this time. The unusual properties of the latter method were to prove useful in some later developments, as we shall see.

One formal development due to Luttinger and Ward(11) in the early sixties is of great interest to solid state theorists. They, in a sense, completed the programme foreshadowed by the Bohm and Pines theory. They showed, using the perturbation method, that electrons near the Fermi surface behave like free (i.e. non-interacting) pseudo-particles in spite of the strong Coulomb interactions; thus they explained why all the earlier calculations of conductivity, Fermi surface properties, etc., had, paradoxically, been successful in spite of their neglect of the interactions, and laid the basis for future calculations of this kind. There is a catch, however. The argument is circular: it shows that solutions with these properties are possible (and are presumably realised in many cases), but it does not show that there are not other solutions with different properties (there are).

Since the beginning of the sixties the pace of real progress has slackened somewhat (although the number of publications in the field has increased enormously). The reason is simple. The early days were much concerned with the development of the technique for the formulation of many-body problems; this phase is now passed. Such technique only enables one to write down the appropriate equations, it does not tell one how to solve them. To make further progress one must find new solutions to these equations corresponding to new physical insights into the problem - a slow and difficult process. It emerged that the Breuckner and plasma oscillation theories were based upon expansions in suitable small parameters. The Breuckner theory is really a low density theory (where pairwise collision processes are dominant) and the plasma oscillation theory is a high density theory (where the simultaneous interactions of many particles have important averaging effects). What we badly need are new and better expansion parameters.

Nevertheless, a number of very important developments have occurred in the intervening years. They have mostly been slanted in a particular direction. The results obtained up to the early sixties had a rather curious aspect. Apart from the notable exception of superconductivity theory, they seemed to show that the particle interactions and correlations, which had been so carefully studied, had little effect! In both the nuclear and the electron gas cases it seemed that one could 'renormalise away' the effects of the interaction and be left with non-interacting pseudo-particles. The subsequent developments that I shall discuss all depart strongly from this position and are concerned with situations in which the interaction and correlation effects have important physical consequences.

One of these developments originated in the Theoretical Physics Division. Its initial aim was to enable one to understand the nature of ferromagnetism in metals (of which more later). It failed to achieve what was hoped for in this direction, but led to bigger things. The problem in question - prompted by the results of neutron scattering experiments - was this: did the magnetic electrons behave as though they formed an electron gas or as though they were localized at the atoms of the ferromagnet? But the question goes deeper: why do some solids behave like a collection of atoms or ions, others as though they contain an electron gas? The answer obtained was that it was a matter of electron correlation: when the correlations are dominant the system behaves like a collection of atoms (a truly important correlation effect). The situation could be studied on the basis of a very simple model (the 'Hubbard Hamiltonian'), the utility of which was such that it has since been applied to a great many other physical situations.

One particularly striking result dropped out of the study of this model(12). In some circumstances the strength of the interaction could determine whether a material was an electrical conductor or insulator. As long ago as 1949 Mott had, with remarkable insight, perceived this possibility and the associated possibility of conductor-insulator (Mott) transitions(13). The new model confirmed the correctness of his view and provided the mathematical apparatus for discussing these transitions.

The critical phenomena occurring at second-order phase transitions have been the subject of intense study in recent years, both theoretical and experimental (the neutron scattering measurements having been of particular value). By the middle sixties these studies had revealed the existence of the 'scaling laws' and 'critical exponents', and a burning question was how to account for these. It turned out that the interaction effects were of dominant importance in this situation also. Kadanoff(14) identified the physical origin of these unusual effects, the loss of the correlation length as a scale length at the critical point (where it diverges). It fell, however, to Wilson(15) to develop the appropriate mathematical apparatus (the renormalisation group theory, again a borrowing from electrodynamics) to study such problems. Here again one has a small parameter, the distance from the critical point, but must adopt a very unusual procedure (repeated renormalisations) to take advantage of it. The theory explains the existence of the scaling laws and their universal nature, and enables one to calculate approximately the critical exponents. It is ideally adapted to the problem of the critical phenomena and similar situations (e.g. the Kondo problem), but is not so obviously suited for the many other applications that have been attempted in recent years. As to why the computational procedures used in conjunction with the renormalisation group method work so well remains a mystery.

I will mention another development rather close to Harwell's interests. I hark back to the problem of ferromagnetism in metals mentioned earlier. About 1970 a new approach to this problem was developed by Evanson, Schrieffer and Wang(16) and by Cyrot(17). It was based upon the rather unusual functional integral technique mentioned above, and provided a kind of unified theory in which one could see how the magnetic electrons could simultaneously exhibit both itinerant (electron gas) and localised behaviour. It is interesting to note that the functional integral technique is the only one presently known to us which could provide the appropriate formalism for this purpose. However, this formalism is very tricky to use and the original formulation of the ferromagnetism problem did not quite correspond to the correct physics of the situation. Recently (1979), however, the method has been revived by Hubbard(18) and Korenman and

Prange(19), using what I believe to be a correct formulation, and has been shown that it might well provide a good description of ferromagnetism in the case of iron (at least).

Finally, I must mention the interesting developments connected with 'low' dimension systems, e.g. the quasi-one-dimensional solids, the liquid helium films, etc. The behaviour of interacting particle systems may depend strongly upon the dimensionality of the space in which they move; the examples mentioned above imitate to some extent many-body problems in one and two dimensions. The one-dimensional case is very unusual: one can sometimes obtain exact solutions! These solutions are all based upon Bethe's 1931 idea. Such exact solutions have been obtained for one-dimensional boson systems, electron gases and even for the one-dimensional Hubbard model. They are of great heuristic value in helping one avoid the traps into which one may fall when using approximate methods (many-body theory is littered with such traps and there have been many casualties). One does not, of course, have phase transitions in one dimension. As a compensation it has been found that in the two-dimensional case one may have transitions of a strikingly unusual nature, dependent on the topology of two dimensions. This remarkable possibility was discovered by Thouless and Kosterlitz(20) in the early seventies; their novel ideas seem now to have been verified by measurements on liquid helium films.

Where does this leave us today? How far have we advanced in 25 years? We have learnt the appropriate language (Green's functions, etc.) to use in discussing many-body systems. Some ancient problems have been solved (why does band theory work so well, what causes superconductivity, why are there different kinds of solids). But we possess no systematic methods of solution. What I have described above is a collection of special solutions appropriate in particular circumstances. Some of these we only half understand. Why do the renormalisation group computational schemes work? Does Luttinger and Ward's result hold for the Hubbard model in the strong interaction case? There are many similar fundamental questions still unanswered. Our understanding of the effects of particle interactions is still limited and we have a long way to go before the subject reaches maturity, let alone senescence.

1. D. Bohm and D. Pines, Phys. Rev. 92, 609 (1953).
2. K.A. Brueckner and C.A. Levinson, Phys. Rev. 97, 1344 (1955).
3. J. Goldstone, Proc. Roy. Soc. A239, 267 (1957).
4. N.F. Mott, Proc. Tenth Solvay Congress, Brussels (1955).
5. H. Frohlich and H. Pelzer, Proc. Phys. Soc. A68, 525 (1955).
6. J. Hubbard, Proc. Phys. Soc. A68, 976 (1955).
7. J. Hubbard, Proc. Roy. Soc. A240, 539 (1957); ibid. A243, 336 (1957).
8. J. Bardeen, L. Cooper and J.R. Schrieffer, Phys. Rev. 108, 1175 (1957).
9. R.L. Stratonovich, Doklady Akad. Nauk S.S.S.R., 115, 1097 (1957) (trans: Soviet Phys. Doklady 2, 416 (1958)).
10. J. Hubbard, Phys. Rev. Lett. 3, 77 (1959).
11. J.M. Luttinger and J.C. Ward, Phys. Rev. 118, 1417 (1960); J.M. Luttinger, Phys. Rev. 121, 942 (1961).
12. J. Hubbard, Proc. Roy. Soc. A276, 238 (1963); ibid. A281, 401 (1964).
13. N.F. Mott, Proc. Phys. Soc. 62, 416 (1949); Phil. Mag. 6, 287 (1961).
14. L.P. Kadanoff, Rev. Mod. Phys. 39, 395 (1967).
15. K.G. Wilson and J.B. Kogut, Physics Reports 12c, 77 (1974).
16. W.E. Evanson, J.R. Schrieffer and S.Q. Wang, J. Appl. Phys. 41, 1199 (1971).
17. M. Cyrot, Phys. Rev. Lett. 25, 871 (1970).

18. J. Hubbard, Phys. Rev. B19, 2626 (1979).

19. R.E. Prange and V. Korenman, Phys. Rev. B19, 4691 (1979).

20. J.M. Kosterlitz and D.J. Thouless, J. Phys. C, 1181 (1973).

Chapter XIV

ATOMIC TRANSPORT IN SOLIDS

A. B. Lidiard

1. Introduction

The subject of diffusion and of atomic migration in solids is a complex one in which the particulars are often as important as the general principles. It is now well established as an integral part of the more general subject of imperfections in solids, since it appears that in all well-studied and well-understood crystalline solids diffusion depends upon the presence of lattice imperfections, particularly the presence of atomic or so-called 'point' defects, i.e. vacant lattice sites (vacancies) and atoms in the interstices of the lattice (interstitials). The early 1950's saw the beginning of this explicit recognition of the integral nature of the study of imperfections in solids(1,2), and in 1960 the first book with this general title appeared(3). The consequent close interactions between research on diffusion and atomic transport with that on, for example, radiation damage, colour centres, mechanical properties, etc., makes it difficult to isolate the currents of progress in diffusion from those in these other topics. Nevertheless, one can discern certain broad developments in diffusion theory and it is these which I shall describe and illustrate in this brief article.

At the root of it all, of course, lies the stimulus provided by the availability of radioisotopes which allowed very precise diffusion measurements to be made under conditions very close to thermodynamic equilibrium. This stimulus was there from the early days of atomic energy. The 1950's and early '60's were a period of fairly rapid innovation in diffusion theory; these developments are summarized in a number of books which appeared several years later(4-7). I believe that one important reason for this burst of theoretical activity was confidence in the basic models of diffusion via vacancy and interstitial defects which derived partly from the fundamental calculations of defect structure by Huntington and Seitz(8) and by Mott and Littleton(9) and partly from the insight into defect structure which it was possible to obtain from experiments on ionic crystals in particular(10).

The 1960's and 1970's have been periods of very notable refinement in experimental technique, one of the objects of which has been the confirmation and application of the principles advanced in the preceding years (e.g. isotope effects, relations between diffusion and ionic conductivity). With the availability of large computers has come a rise in the importance of basic calculations of defect structure and energies which have enabled us in the 70's to understand complex structures which didn't fit in with the simpler models used for the earlier diffusion theories (e.g. the Willis structures in UO_{2+x}). The results of these calculations are now being increasingly incorporated into diffusion theories and analyses(11). Amongst these computer-based calculations are those using Monte Carlo and molecular dynamics techniques, which enable us to describe systems for which the simpler, essentially static models are inadequate. All these methods will undoubtedly become more important in the future. Even so the simpler analytic theories remain to be unified.

This article attempts to illustrate these broad developments in more detail. The next section deals with models of defects - although only briefly since this is the subject of Chapter XVII by Michael Norgett. Section 3 then discusses the statistical mechanics of defect populations, while Section 4 deals with theories of the diffusion and migration of atoms by the action of these defects. Unfortunately, there is space to deal only with the fundamentals but not with applications such as precipitation or nucleation theory. Ron Bullough's article on Radiation Damage in Metals does, however, deal with one such application, namely defect movements under irradiation.

2. Models

That atoms and ions can move through solids has been known for a long time. Before we can begin to provide quantitative theories of these movements we need ideas about how these movements occur; models in other words. By the beginning of the period of this review the idea that these movements were accomplished by the action of lattice imperfections - vacant lattice sites and interstitial atoms or ions - was well established, although the idea that pairs or larger groups of atoms could exchange places directly by a thermally activated co-operative movement was also still being explored(12). These imperfections allow the atoms of the solid to move to an extent proportional to their concentration. Thus the atoms next to a vacancy may, if sufficiently 'activated' by thermal motion, jump into the vacant position. Likewise an interstitial atom may jump directly from one interstitial site to another by thermal activation. Or it may move by pushing a neighbouring atom off its lattice site into a new interstitial position and itself taking the lattice site so vacated. And one can envisage other possibilities.

However, despite the computational difficulties, solid state theory had already begun to guide the choice of models. Huntington and Seitz(8) had studied the energetics of the vacancy and the interstitial atom in Cu and obtained several results which had a considerable influence. In particular, they showed that the energy of formation of the vacancy was only about one-third of that of the interstitial, so that vacancies would be the dominant thermally produced defect. Furthermore, the calculated magnitudes of the vacancy formation and activation energies were such as to explain the experimentally observed diffusion rates. Thirdly the relaxation of the lattice around the vacant site was quite small and no radical re-arrangement of the local co-ordination was indicated. These results were assumed to hold for the other noble metals and indeed for close-packed metals generally. They gave a confidence to the theory of diffusion in metals which allowed it to flourish in the 1950's and 1960's. A similar confidence was imparted to the theory of diffusion and ionic transport in ionic crystals by the even earlier calculations of Mott and Littleton(9) for the alkali halides. These again showed that the dominant thermally produced defects were vacancies (Schottky defects, i.e. anion and cation vacancies together) and that the lattice co-ordination around the vacancies was not drastically perturbed. This confidence that solid state diffusion was to be understood in terms of the movements of distinct defects, particularly vacancies, also had the additional important consequence of linking the study of diffusion to other branches of solid state physics, chemistry and metallurgy concerned with imperfections, and especially with the new subject of radiation damage stimulated by nuclear reactor technology. That these links were well established at the beginning of our period may be seen from two reviews of Seitz from that time(1,13).

But to return to diffusion theory. By 1954 the atomistic calculations had been enlarged to include the energies of the interaction of solute atoms and ions with vacancies, finding them to be weak and of short range in a metal(14) and rather larger and of long-range in ionic crystals(15,16). Diffusion theory had, in fact, all that was needed to set up definite models for many of the statistical analyses of the next 25 years*. These analyses were made particularly actively in the next dozen years or so (see section 4): they were drawn together in the books of Adda and Philibert(6) and of Manning(7). But the atomistic calculations had also flourished in this time and in 1966 the first conference purely devoted to such calculations was held under the auspices of the National Bureau of Standards in Washington(18). Experimental studies were disclosing or implying some complex and unexpected defect clusters (e.g. those in UO_{2+x} and in $Fe_{1-x}O$) as well as showing up those which were expected. It was necessary to return to the atomistic calculations to understand these. As Chapter XVII by Michael Norgett shows, very considerable progress has been made since that first conference in 1966; for example, we now understand well the origin of the clusters in $Fe_{1-x}O$ and in UO_{2+x} and related substances. The results of these atomistic calculations are now increasingly calculations are now increasingly used to guide theories of diffusion and transport in very specific and sophisticated ways.

In keeping with the rather rigid atomic structure around most defects these atomistic calculations have generally been made on static lattices - which can be justified within the framework of the quasi-harmonic approximation. But the late 70's have seen one important example which shows

*This is certainly true of ionic crystals and most metals. Some alloy systems have, however, been found subsequently to exhibit very rapid solute diffusion (e.g. a number of elements in Pb and in Sn). These were investigated notably in the late 60's and in the 70's and special defect models devised(17). At the present time there are still some uncertainties about the correct defect models to use for semiconductors.

that there are limitations to this approach; for this we must use a fully dynamical approach such as that offered by the 'molecular dynamics' method of computer simulation. The example is the class of compounds having the fluorite (CaF_2) structure. These materials were extensively studied in the 60's and appeared to be analogous to the alkali halides, the differences following naturally from their different lattice structure. When we wrote the book 'Crystals with the Fluorite Structure' under Bill Hayes' editorial leadership at the beginning of the 70's there were just a few anomalies, most notably a rather little known indication of a high temperature specific heat peak. Soon afterwards it was shown that at temperatures above this peak several of these substances showed high ionic conductivities and high rates of anion self-diffusion. The molecular dynamics calculations show that this high temperature behaviour is quite different from the low temperature behaviour. In particular, the local structure of the defects appears to change(18). In this class of material we therefore lose the simplicity of the early and well-established models of vacancies and interstitials when we go to temperatures close to the melting point.

3. Statistical Mechanics

The theory of atomic transport by lattice imperfections in effect separates observable transport or diffusion coefficients into products of two factors, one proportional to the local density of imperfections and the other involving the frequencies of movement of the atom by the imperfections. In this section we discuss the density of these defects.

The densities of defects such as vacancies and interstitials may be obtained by application of the methods of statistical mechanics. For the simplest examples of dilute systems of non-interacting defects the results were well established in the scientific literature even by the outbreak of the second World War. The concentration of such defects will depend exponentially on temperature through a Boltzmann factor which contains their formation energy. Indeed, this result is so obvious and so well known that it may seem hardly worthy of mention. But even this result will only be valid as long as the temperatures, and thus the defect mobilities, are high enough to ensure thermodynamic equilibrium.

The qualification is important historically. For a series of elegant experiments carried out in the 1950's (notably by Koehler et al at the University of Illinois) showed that it was possible to cool hot wires and foils of metals sufficiently rapidly that one could retain the vacancy population established at the original high temperature. Subsequent annealing of these specimens at intermediate temperatures allowed the vacancy mobilities to be determined. Thus it was possible (a) to find the density of vacancies as a function of temperature, (b) to determine their mobility as a function of temperature and (c) to show that these defects were responsible for the diffusion of atoms in these metals.

Alongside these quenching experiments were the outstandingly precise 'Simmons-Balluffi' X-ray and dilatometric experiments, which provided direct demonstrations of the presence of vacancies in metals and dilute alloys and gave absolute measures of their concentration.

These two types of experiment thus confirmed the vacancy model of diffusion in the most direct ways possible and thereby greatly strengthened the theory of diffusion, especially in metals. A somewhat analogous strengthening of the theory of ionic transport in the alkali and silver halides had been provided earlier by the discovery that one could control the density of vacancies in such crystals by the addition of small concentrations of foreign ions of a valency different from that of the corresponding host ion* (e.g. Cd^{2+} in NaCl).

Thus the elementary theory of non-interacting defects was established before the beginning of our period and well confirmed experimentally in several distinct classes of solid (metals, solid rare gases, ionic crystals) in the years following. But how was theory to deal with interactions between these defects? We may distinguish interactions at short range and at long range. In a metal, for example, two vacancies may attract one another with a 'binding energy' of roughly 0.1 eV when they are nearest neighbours but have a quite negligible interaction at greater separations(20). In an ionic crystal an anion vacancy and a cation vacancy will also bind together when close to one another (even as strongly as 1 eV) but they will also have a long-range Coulombic interaction at all distances.

When only short-range interactions are significant it is natural to treat the close pairs as distinct complex or quasi-molecular defects. In first approximation their concentration will be obtained from the corresponding 'mass-action' equation. In complex systems (e.g. non-stoichiometric compounds) there will generally be several of these mass-action equations to be satisfied simultaneously. Although there were no

*For such ions John Sykes proposed the adjective 'aliovalent' ('of another valency', from the Latin 'alius'), a word which I put into circulation via my review of ionic conductivity in the Handbuch der Physik(10). It is now so well established as occasionally to be mis-spelt!

very challenging fundamental problems for theory here, there is no doubt that the technique developed by Kroger and Vink(21) for mapping out the nature of the solutions to such sets of equations was very valuable, especially for compound semiconductors for which it is still used.

In ionic crystals and in semiconductors where long-range Coulombic interactions may also be present it seemed sensible to appeal to Debye-Hückel theory while again retaining the idea of pairs or larger clusters of defects as distinct entities(10,22). This was no more than had already been done in liquid electrolyte theory. This idea provided a theoretical framework for the refined representation of ionic transport measurements and served the subject well throughout the 1950's and 60's. In the last ten years, however, there have been several studies, most notably by Allnatt and co-workers, to determine the limits to this application of Debye-Hückel theory(23). These limits occur for conditions which are within the ranges studied experimentally in, for example, the Ag-halides and in doped fluorite compounds. The implications for transport, as distinct from thermodynamic properties, have yet to be fully worked out, however.

So far we have discussed dilute defect systems. But there are many systems where large concentrations of defects or impurities lead to special ordering or co-operative phenomena. Examples are provided by non-stoichiometric oxides of transition metals, fluorite solid solutions, various fast ion conductors, etc. For these the well tried, basic approximations of the theory of order-disorder phenomena, regular solutions, etc. (e.g. the molecular field approximation) are frequently used. Uncertainty over important details of the models means that the refinements in order-disorder theory which have occurred in the past 25 years are not often used in these applications. Indeed, many of these are quite intuitive and may even be bad phenomenologically. For example theories of the 'fast ion conductivity' of the fluorite compounds at high temperatures[24], which lead to a rapid and co-operative increase in the number of Frenkel defects at high temperatures, may be correct qualitatively but incorrect quantitatively in indicating that as many as one-half of all the anions are in interstitial positions at these temperatures. For at least some of the fluorite compounds there is both experimental and theoretical evidence that the degree of disorder at high temperatures is only a few per cent. The theoretical evidence comes from very complete molecular dynamics calculations[19]. The fluorite compounds would seem to provide a very satisfactory ground for the future development of ideas of co-operative defect transitions; they can be modelled satisfactorily, their high temperature properties are accessible experimentally and a great deal is already known of their low temperature defect properties.

4. General Theories of Diffusion

So far we have considered the models of defects responsible for atomic diffusion in solids and the concentrations of these defects. In this section we consider aspects of the theory by which these models are related to observable diffusion and other transport coefficients. We begin with absolute rate theories of defect and atomic mobility (e.g. Eyring, Vineyard). We then discuss (i) theories of relaxation processes (e.g. dielectric and anelastic relaxation), (ii) kinetic theories, (iii) random walk theories and (iv) the development of the irreversible thermodynamics formalism in the light of the advances achieved by means of these other more particular methods.

4.1 Absolute Rate Theories

These provide the basic (Arrhenius) expression for the frequency of the diffusive jump of an atom from one lattice site to another. They derive from the theory of chemical kinetics(25) and, for classical systems in particular, from the work of Eyring(26). The derivations of the atomic jump frequency were tidied up by Wert in 1950 and especially by Vineyard(27) in 1957. The Vineyard expression remains the valid general expression of the essential idea of classical absolute rate theory. Within the framework of the quasi-harmonic approximation it tells us how to calculate activation energies and entropies and isotopic effects (i.e dependence of jump rates on the mass of either the moving atom or a neighbouring atom). The result for the activation energy is very simple: the activation energy is just that in the corresponding static lattice. As a result atomistic calculations of activation energies are almost invariably done this way. By contrast there have been very few calculations of defect entropies(28): these require an evaluation of vibrational frequencies but the prescription is clear-cut and the tools are available in the form of the HADES and PLUTO programs (see chapter XVII). The calculation of isotope effects also requires the calculation of a frequency - that of the decomposition mode at the saddle point (which is a purely imaginary number). Again the presciption is clear-cut and the tools are now available. But there have been very few such calculations. That by Ron Bullough, Norman March, Roy Perrin et al was on vacancies in Cu, Na and NaCl(29), but that is now almost ten years old. In view of the considerable experimental interest in isotope effects[30] this situation ought not to continue for much longer.

The Vineyard formulation is based on classical stastical mechanics. As such it should not be valid at low temperatures, i.e. at temperatures where we know that the thermal motions must be described quantum mechanically. Nevertheless, atomic movements at very low temperatures, e.g. those of interstitial atoms in close-packed metals, take place with a frequency which often depends upon temperature in the classical Arrhenius way. The general theory of such movements does not yet appear to have been developed in a satisfactory way. A particular exception is the quantum theory of the diffusion of very light elements[31], hydrogen and helium, developed originally by Pete Flynn and Marshall Stoneham during the latter's sabbatical visit to the University of Illinois in 1969. This theory has recently seen another application - to the description of the motion of 'thermalized' muonium in solids.

4.2 Relaxation Processes

One of the simpler tests of theories of atomic jump frequencies is provided by measurements of the rates of certain anelastic and dielectric relaxation processes. Perhaps the simplest example is that of the so-called Snoek relaxation associated with the slight tetragonal distortion around interstitial impurities (e.g. C and N) in b.c.c. metals where a single jump frequency is involved. Certainly this example was historically significant[12]. In other cases, e.g. those involving pairs or larger clusters of defects and impurity atoms, several distinct jump frequencies may enter into the (several) characteristic relaxation times. As often, the alkali halides provided a convenient testing ground for simple ideas. In 1955 I discussed[32] the dielectric relaxation of impurity-vacancy pairs in ionic crystals and thereby introduced what is now called the 'five-frequency model' which has served as the model for much subsequent diffusion theory, not only of the alkali halides but for f.c.c. metals as well. This discussion, however, used elementary methods. Symmetry analysis was later employed for this and related models for other crystal structures by Alan Franklin[32] and by others. These methods for determining relaxation modes and their symmetry, and thus the fields they couple to were extensively developed by Nowick and are well described in his reviews[34] from the 1960's. But the number of systems which can be said to be fully analysed, i.e. all relevant jump frequencies determined, is still small even today.

4.3 Kinetic Theories

Many of the earlier elementary discussions of solid state diffusion were really of this type, e.g. those which counted up the numbers of atoms crossing from one lattice plane to another in order to determine the net fluxes of atoms. Seitz set down such a theory for the Kirkendall effect in 1950[35] but it was hardly possible at that time to be very explicit for a concentrated alloy. Rather we needed to develop ideas on dilute alloys first. In 1955 I presented a kinetic theory of such a dilute alloy using the five-frequency model of diffusion via vacancies[36]. Actually at the time I was more concerned with solute diffusion in ionic crystals than in metals, since in these materials more striking effects could be expected (e.g. the rapid dependence of diffusion coefficient upon solute concentration, first glimpsed experimentally in the Ag-halides in 1951 and later very precisely determined during the 1960's and 70's by Fredericks et al in the alkali halides[37]). But the approach was general and the particular equations could be used for any crystal with the f.c.c. structure or for diffusion on any f.c.c. sub-lattice in a compound.

One of its immediate consequences was to draw attention to what is now known as the 'correlation factor' in atomic diffusion and to provide an approximate form for it in terms of the various vacancy jump frequencies in the vicinity of a solute atom. This factor expresses the fact that when an atom diffuses by the action of a vacancy (or other separate defect) its successive movements are statistically correlated with one another; for example, an atom which has just jumped into a vacancy is more likely on its next jump to return to its former position than to move to other sites - simply because the atom can move only to a site which is vacant. Actually attention was first drawn to this feature of diffusion via vacancies in two earlier papers on solute diffusion - the first by Johnson[38] in 1939 and the second by Wyllie[39] in 1947 - which were, however, either incorrect or incomplete in their analysis. The idea was again discussed by Bardeen and Herring[4] in 1952 in connection with self-diffusion (i.e. where the 'solute' is chemically indistinguishable from the solvent).

Other consequences of the explicit expressions provided by this kinetic formulation followed later and included (i) expressions for ionic mobilities and generalizations of the Nernst-Einstein relation[5], (ii) a theory of the influence of solutes upon self-diffusion rates[41], (iii) a description of the motion of solute atoms induced by a flux of vacancies[42,43], (iv) interpretations of the phenomenological coefficients, L_{ij}, appearing in the thermodynamics of irreversible processes(5).

The same approach has been used subsequently by others for other diffusion problems(44). These include diffusion in a temperature gradient (Soret effect) as well as isothermal diffusion and interstitial and dissociative alloys as well as substitutional alloys.

There are some disadvantages to the method. One is its algebraic complexity and the absence of a clear formal structure related to the symmetry of the problem. The second is that it gives approximate results but without a very clear formal sequence of successive approximations. Nevertheless, I believe that the structure of the kinetic theory and its relation to other approaches, especially the random-walk method (see below), can be made more apparent; indeed the work of Manning(7) and of Yoshida(44c) points the way. There is a close relationship to the theory of relaxation modes (4.2 above) waiting to be picked out too. These extensions are worth pursuing on account of the ease with which this kinetic approach can be put into correspondence with irreversible thermodynamics which in turn allows further applications to be made(37,43).

4.4 Random Walk Theories

By such theories we mean those that derive from the Einstein equation which expresses the diffusion coefficient, D, in terms of the mean square displacement of the diffusing atoms in a time, t. These theories evaluate the mean square displacement in terms of the parameters of the model (defect concentration, jump frequencies) by means of statistical arguments. This approach leads naturally to the appearance of the correlation factor in the diffusion coefficient. Indeed that was one principal result of the 1952 Bardeen-Herring paper referred to above(4). Alan LeClaire and I then saw how the result which I had obtained for solute diffusion by the kinetic method could be confirmed by the random-walk method(45). At the same time Compaan and Haven (at Eindhoven) were evaluating correlation factors for self-diffusion via vacancies(46) and then for the interstitialcy or replacement mechanism(46). Later work by John Manning(7) and Bob Howard(54) (at the N.B.S. Washington) resulted in the development of several mathematically elegant ways of calculating these correlation factors for all cases of practical interest; these methods were nicely brought together and reviewed by Alan LeClaire in 1970(47). Others have been added since(55).

But why should these correlation factors be of interest? For a given lattice they depend upon the mechanism of diffusion (vacancy or interstitialcy). For solute diffusion via vacancies they depend on the relative rates at which solute atoms and solvent atoms jump into vacancies. For reasons such as these it was desirable to determine these factors experimentally as well as to calculate them. Ionic conductors such as NaCl offered one possibility via a comparison of self-diffusion coefficient and ionic conductivity (to give the Nernst-Einstein ratio) as noted early on by Haven(48). This possibility was realized (somewhat unexpectedly) in a strikingly clear way in AgCl by Dale Compton at the University of Illinois and explained by Charles McCombie and myself as a manifestation of Ag^+ ion diffusion by the interstitialcy (replacement) mechanism(49). One had known that Ag^+ Frenkel defects were the dominant defects in AgCl and AgBr but up to that time had thought that the Ag^+ interstitials jumped directly from one interstitial site to another. The calculations of Huntington and Seitz[8] had pointed to the possibility of the intersitialcy (replacement) mechanism in Cu but this was the first time the occurrence of such movements was directly demonstrated experimentally. The value of accurate comparisons of ionic mobility and diffusion was thereby established and has been much relied on since. Generalisations for solute ions were provided later, notably by Manning[7] and in the thermodynamic formalism by Bob Howard and myself[5].

But, while one can determine ion mobilities in ionic conductors and semiconductors, in metals a different approach is needed if one is to determine this correlation factor. Schoen[5] in 1958 initiated one very fruitful approach, namely that of isotope effect measurements, by observing that the relative difference in diffusion coefficient of two isotopes of the same element is simply related to the product of the correlation factor with a term which is the relative difference in the frequency of exchange with the vacancy, and this term, by elementary rate theory, he set equal to the relative difference in the inverse square root of the mass number of the isotopes. The simplicity of this result made it very important. Schoen had derived it from our approximate expression for the correlation factor in a f.c.c. lattice but Tharmalingam and I were soon able to show that it applied rather generally to a range of lattice symmetries. The range of conditions has been subsequently (1971) found to be even wider[51]. However, because these isotope effects are so small, very precise measurements are required to take advantage of these theoretical results. Techniques were developed at Harwell in the 1960's, principally by Les Barr, and at much the same time by Norman Peterson at Argonne. Those at Argonne still fluorish; for a review see the article by Peterson(52).

The universality of the method makes it very valuable. But there is an important limitation. That is the assumption that only the jump frequency of the diffusing atom depends upon the isotopic mass of the atom and that it does so as the inverse root mass. The Vineyard formulation of the absolute rate theory(27) shows that these assumptions are not necessarily accurate, although it allows us to calculate the corrections. The result is that the isotope effect yields not simply the correlation factor alone but its product with another factor which is dynamical, rather than geometrical, in origin. As we have already remarked (in section 4.1 above) there have been few attempts to calculate it yet and experimentalists have therefore had to get along with empirical values as best they could(52). In this they have had assistance from the results on ionic mobilities mentioned above and from the results of quite different experiments on the influence of solutes upon both solvent and solute diffusion rates. For the interpretation of these, approximate theories were again useful(41), although superseded by more accurate ones(53) later.

4.5 Thermodynamics of Irreversible Processes

By the beginning of our period, several important works(56) on this subject had recently appeared. These contained three important basic points. The first was the representation of the phenomenological relations between the 'flows', J_i, (of heat, matter, electricity, etc.) and the 'forces', X_j, causing these flows as

$$J_i = \sum_j L_{ij} X_j \ . \qquad (1)$$

The second was the result that the rate of production of entropy in the system, $\dot{S}$, was given by

$$T\dot{S} = \sum_i J_i X_i \ . \tag{2}$$

The third was the Onsager relation between the off-diagonal coefficients, namely

$$L_{ij} = L_{ji} \ . \tag{3}$$

Obviously eqns. (2) and (3) imply that only certain ways of representing the forces (gradients of temperature, concentration, electric potential, etc.) are admissable. Unfortunately, de Groot's book(56d), while an important and popular work, did not distinguish clearly enough between the sufficient and the necessary conditions for (3) to hold*, and this led to trouble later.

This representation of solid state diffusion, with the inclusion of vacancies as a distinct species, was used in the early 50's by Bardeen and Herring(40) and by Le Claire(54). However, for the Onsager relations (3) to be useful one needed a greater understanding of these phenomenological coefficients in relation to the underlying atomic movements. A beginning was made by Bardeen and Herring who showed, for the special case of self-diffusion by tracers (A* in A), that the off-diagonal coefficient, L_{AA*}, was directly proportional to 1-f, f being the correlation factor (4.4 above). After the development of the kinetic theory of solute diffusion (section 4.3 above) it was possible to go much further and Bob Howard and I

*The matter is handled more satisfactorily in the later book by de Groot and Mazur(57) but has still come in for some fierce criticism(58).

extended the application and analysis of this formalism for solid state transport via vacancies fairly substantially in our 1964 review(5). Very similar ideas were in the minds of Adda and Philibert at that time and both the phenomenology and the interpretation of it provided by the kinetic theory figures prominently in their volumes on diffusion(6) which appeared in 1966. That this formalism is of practical value and convenience has also been shown by subsequent applications to (i) solute drift phenomena occasioned by vacancy fluxes(43) and (ii) the complexities of the diffusion of aliovalent solutes in ionic crystals(37). However, it is disappointing that more progress has not been made in the determination of the L_{ij} coefficients for other defect systems(44f). Despite the more fundamental formulation offered by Allnatt, the straightforward pair approximation type of kinetic theory still seems the most immediate way to obtain these coefficients. Furthermore, it is not obvious that we can obtain all the L_{ij} coefficients for a given system from the random walk method; and at present it appears that we cannot. This emphasizes the desirability of refining and generalizing the kinetic theory - as already noted above (section 4.2).

An important pair of conjugate couplings represented in the theory of non-equilibrium thermodynamics are those between a temperature gradient and a matter flow and between a concentration gradient and a heat flow. Analyses of the resulting phenomena such as the Soret effect and ionic thermopower(5,61) drew attention in the 1950's and 60's to the heat-of-transport parameter, q^*, for defects. While this parameter has thereby been successfully determined experimentally for many systems there has been almost no theory of it - as distinct from rather barren intuitive discussions - until very recently. Thanks to the work of Mike Gillan, however, we now have a method for the calculation of q^* from a potential model of the solid(62). This should enable us to obtain heats of transport with as much confidence as we presently obtain energies of defect formation and activation. These calculations represent an important task for the next few years.

5. Conclusion

The study of diffusion and atomic migration is part of the wider study of imperfections in crystalline solids. This was recognised already by the early 1950's. In the next 25 years diffusion and transport studies added enormously to our knowledge of the intimate details of defect structure, energies and movements. This was obtained not merely from measurements with radioisotopes but by many other techniques as well, e.g. dielectric and mechanical relaxation, nuclear magnetic relaxation, ionic conductivity and thermocurrents, electron paramagnetic resonance, etc. Corresponding developments in the theory of defects and diffusion were, however, necessary to translate these measurements into knowledge and understanding. It is these theoretical developments which have been surveyed here. As a result we can see some of the developments which will occur in the next few years. The newer methods of computer simulation - Monte Carlo and molecular dynamics - will undoubtedly play an increasingly important role in the study of strongly interacting systems (e.g the fluorites at high temperatures) and in the evaluation of heats of transport (by Gillan's method). At the same time we can expect to see some basic unification of analytic theories as well as their extension to a growing diversity of particular systems. In particular, as the effects of radiation in fast reactor environments are sufficiently great as to cause significant diffusion (or rather drift) effects there will be further developments relating to interstitial-induced diffusion parallel to those described here relating to vacancy-induced diffusion[44e]. There will undoubtedly be many other theoretical developments besides these few examples. However, I am sure that diffusion theory will retain many of the characteristics visible during the past 25 years. It will continue to provide challenges to our physical understanding and to our mathematical and computational skills and insights. It will continue very close to experimental studies at many points but not always so close that it will be without its controversies, mistakes and eccentric aspects (but not, of course, at Harwell!).

1. See Imperfections in Nearly Perfect Crystals, Eds. W. Shockley et al (Wiley, New York, 1952) and especially the article by F. Seitz (p.3).

2. See Defects in Crystalline Solids, being the proceedings of the conference held at Bristol in July 1954 (Physical Society, London, 1955).

3. H.G. Van Beuren, Imperfections in Crystals (Amsterdam, North Holland, 1960).

4. P.G. Shewmon, Diffusion in Solids (McGraw Hill, New York, 1963).

5. R.E. Howard and A.B. Lidiard, Rep. Prog. Phys. 27, 161 (1964).

6. Y. Adda and J. Philibert, La Diffusion dans les Solides (Presses Universitaires de France, Paris, 1966) 2 vols.

7. J.R. Manning, Diffusion Kinetics for Atoms in Solids (Van Nostrand, Princeton, 1968).

8. H.B. Huntington and F. Seitz, Phys. Rev. 61, 315 (1942) and H.B. Huntingtcn, Ibid p.325.

9. N.F. Mott and M.J. Littleton, Trans. Far. Soc. 34, 485 (1938).

10. A.B. Lidiard, Handbuch der Physik 20, 246 (1957).

11. See, e.g., the proceedings of the 3rd Europhysics Conference on Lattice Defects in Ionic Crystals held at Canterbury in 1979 (J. Phys. (Paris) in press).

12. C. Zener, Ref. 1, p.289.

13. F. Seitz, Rev. Mod. Phys. 26, 7 (1954).

14. D. Lazarus, Phys. Rev. 93, 973 (1954). This picture was strengthened not long afterwards by L.C.R. Alfred and N.H. March, Phil. Mag. 2, 985 (1957).

15. J.R. Reitz and J.L. Gammel, J. Chem. Phys. 19, 894 (1951).

16. F. Bassani and F.G. Fumi, Nuovo Cim. 11, 274 (1954).

17. For a review of the current position with these alloys see W.K. Warburton and D. Turnbull in Diffusion in Solids - Recent Developments, Eds. A.S. Nowick and J.J. Burton (Academic Press, New York, 1975) p.172.

18. Calculation of the Properties of Vacancies and Interstitials, Ed. A.D. Franklin, N.B.S. Misc. Publ. No. 287 (1967).

19. M. Dixon and M. Gillan, AERE Reports TP.794 and TP.798 (1979): J. Phys. C. (to be published).

20. A. Seeger and H. Bross, Z. Phys. 145, 161 (1956).

21. F.A. Kröger and H.J. Vink, Solid State Physics 3, 307 (1956).

22. H. Reiss, C.S. Fuller and F.J. Morin, Bell System Tech. J. 35, 535 (1956).

23. See, e.g., A.R. Allnatt and P.S. Yuen, J. Phys. C 8, 2199 (1975) and A.R. Allnatt, ref. 11.

24. See, e.g., B.A. Huberman, Phys. Rev. Letts. 32, 1000 (1974); M.J. Rice, S. Strassler and G.A. Toombs, Ibid 32, 596 (1974).

25. C.N. Hinshelwood, The Kinetics of Chemical Change (Oxford University Press, 1940).

26. S. Glasstone, K.J. Laidler and H. Eyring, The Theory of Rate Processes (McGraw Hill, New York, 1941).

27. G.H. Vineyard, J. Phys. Chem. Solids 3, 121 (1957).

28. See J. Govindarajan, P.W.M. Jacobs and M.A. Nerenberg, J. Phys. C 9, 3911 (1976) and 10, 1809 (1977) and references cited there.

29. R.C. Brown et al, Z. Naturforsch 26a, 77 (1971) and Phil. Mag. 23, 555 (1971).

30. N.L. Peterson, ref. 17, p.115.

31. A.M. Stoneham, Berichte der Bunsen Gesellschaft für Physikalische Chemie 76, 816 (1972).

32. A.B. Lidiard, ref. 2, p.283.

33. A.D. Franklin, A. Shorb and J.B. Wachtman, J. Res. National Bureau of Standards, 68A, 425 (1964).

34. A.S. Nowick and W.R. Heller, Adv. Phys. 12, 251 (1963) and 14, 101 (1965); A.S. Nowick, Ibid 16, 1 (1967).

35. F. Seitz, Acta Cryst. 3, 355 (1950).

36. A.B. Lidiard, Phil. Mag. 46, 1218 (1955). For a correction to this paper see R.E. Howard and A.B. Lidiard, J. Phys. Soc. Japan 18, Suppl. II, 197 (1963).

37. For a current review see W.J. Fredericks, Ref. 17, p. 381.

38. R.P. Johnson, Phys. Rev. 56, 814 (1939).

39. G. Wyllie, Proc. Phys. Soc. 59, 694 (1947).

40. J. Bardeen and C. Herring, ref. 1, p. 261. See especially Appendix A, but note the presence of errors in eqns. (A.2) and (A.9) which invalidate the numerical results.

41. A.B. Lidiard, Phil. Mag. 5, 1171 (1960).

42. R.E. Howard and A.B. Lidiard, Phil. Mag. 11, 1179 (1965).

43. T.R. Anthony, ref. 17, p.353.

44. These include:
(a) H. Reiss, Phys. Rev. 113, 1445 (1959).
(b) R.E. Howard and J.R. Manning, J. Chem. Phys. 36, 910 (1962).
(c) M. Yoshida, Japan J. Appl. Phys. 10, 702 (1971).
(d) R.A. McKee, Phys. Rev. B13, 635 (1976) and B15, 5612 (1977).
(e) A. Barbu, Contribution à l'Etude des Changements de Phase sous Irradiation Rapport CEA-R-4936 (1979).
(f) A.B. Lidiard and R.A. McKee, ref. 11.

45. A.D. LeClaire and A.B. Lidiard, Phil. Mag. 1, 518 (1956).

46. K. Compaan and Y. Haven, Trans. Faraday Soc. 52, 786 (1956) and 54, 1498 (1958).

47. A.D. LeClaire in Physical Chemistry - an Advanced Treatise (Academic Press, New York 1970) Vol. X, p.261.

48. Y. Haven, ref. 2, p.261.

49. C.W.McCombie and A.B. Lidiard, Phys. Rev. 101, 1210 (1956).

50. A.H. Schoen, Phys. Rev. Letts. 1, 138 (1958).

51. H. Bakker, Phys. Stat. Solidi 44, 369 (1971).

52. N. Peterson, ref. 17, p.115.

53. See for example the review by A.D. LeClaire, J. Nucl. Mat. 69 and 70, 70 (1978).

54. R.E. Howard, Phys. Rev. 144, 650 (1966).

55. G.E. Murch and R.J. Thorn, Phil. Mag. A39, 673 (1979).

56. (a) S.R. de Groot, L'Effet Soret (North Holland, Amsterdam 1945).
(b) I. Prigogine, Etude Thermodynamique des Phénomènes Irreversibles (Desoer, Liège 1947).
(c) K.G. Denbigh, The Thermodynamics of the Steady State, (Methuen, London 1951).
(d) S.R. de Groot, Thermodynamics of Irreversible Processes (North Holland, Amsterdam 1952).

57. S.R. de Groot and P. Mazur, Non-Equilibrium Thermodynamics (North Holland, Amsterdam 1962).

58. C. Truesdell, Rational Mechanics (McGraw Hill, New York, 1969).

59. A.D. Le Claire, Prog. Metal Phys. 4, 265 (1953).

60. A.R. Allnatt, J. Chem. Phys. 43, 1855 (1965); A.R. Allnatt and L.A. Rowley, Ibid 53, 3217 and 3232 (1970).

61. A.R. Allnatt and A.V. Chadwick, Chem. Rev. (1967) p.681.

62. M.J. Gillan, J. Phys. C 10, 1641 and 3051 (1977); M.J. Gillan and M.W. Finnis, Ibid 11, 4469 (1978).

Chapter XV

RADIATION DAMAGE IN METALS

R. Bullough

Research on radiation damage in metals has always been and continues to be an important experimental and theoretical activity at Harwell. In the last 20 years or so this research at Harwell and elsewhere has led to a huge literature on the subject which it is quite impossible to review adequately in the present brief report. It must suffice to be highly selective and to emphasize only some of the important developments. I shall deal particularly with those where progress has been due to, or has occurred in conjunction with, theoretical studies at Harwell.

There is little doubt that the discovery of void swelling in the stainless steel fuel element cladding in the Dounraey Fast Reactor (D.F.R.) in 1967(1) changed the emphasis in radiation damage research. Before this discovery it was generally thought that in non-fissile metals at temperatures where the vacancies and interstitials are mobile any point-defect aggregates that form would be planar and thus that changes in the shape of the specimen would only result if the metal was crystallographically anisotropic (non-cubic). The observation of three-dimensional vacancy aggregates in the (cubic) steel cladding was therefore a surprise. The understanding of this phenomenon thus called for a detailed study of the influence of the entire microstructure and of any gaseous impurities on the formation and growth of the point-defect aggregates. Whilst swelling and other associated phenomena such as growth and irradiation creep are still not completely understood and we are a long way from being able to design a damage-resistant alloy without experimental testing, we do now have a comprehensive understanding of the many factors that influence the response of metals and alloys to radiation. Because of these relatively successful theoretical developments, we shall tend to emphasize progress since 1967. This date also coincides with rapid developments in our ability to observe and interpret the images of point-defect clusters and other extended defects in the transmission electron microscope (T.E.M.); the progress in understanding stems largely from collaborative research between the theory and T.E.M. studies.

The properties of point-defect clusters had been studied extensively before void swelling was first observed and much of this work provided a useful background of understanding against which to tackle the swelling problem. Thus in 1959 Greenwood, Foreman and Rimmer(2) provided an important analysis of the role of vacancies and dislocations in the nucleation and growth of gas bubbles in fissile materials. The growth of the bubbles in fissile metals like uranium at high temperatures is driven by the accumulation of fission gas to high pressures within the cavities, with the consequent forced provision of 'thermal' vacancies from nearby vacancy sources, such as dislocations and grain boundaries. Although their simple model of bubble nucleation has been largely superseded by recent, more powerful, theories of nucleation, their analysis of bubble growth remains basically correct and was something of a milestone because it drew attention, for the first time, to the importance of the microstructure of the metal (e.g. purity, dislocation content, grain size and shape, density and size of precipitates). In addition, they pointed out that, at lower temperatures where the rate of thermal emission of vacancies is low, gas bubbles might grow to become voids if the dislocations provided a preferential sink for interstitials compared with vacancies. We now know that such void growth does happen - and for the reasons they suggested some ten years before its observation.

The irradiation growth of uranium (i.e. its extension in one dimension) was postulated[3] to arise from the separation of vacancy and interstitial loops on to different crystallographic planes. The growth of non-fissile anisotropic metals such as zirconium and Zircaloy was explained similarly[4]. This model of uranium growth was subsequently confirmed by theoretical calculations[5] using anistropic elasticity theory to describe the interactions of dislocation loops. Elastic energy calculations for dislocation loops were also made earlier by Bullough and Foreman[6] to explain the morphology and orientation of loops in f.c.c. materials and by Eyre and Bullough[7] to explain the properties of loops in b.c.c. materials. The b.c.c. work was later (1968) amplified by Bullough and Perrin[8], who developed computer simulation techniques to follow the nucleation of interstitial clusters. They were able to demonstrate, for the first time, the transition from an aggregate of a few point defects to a glissile dislocation loop that rotated on its glide prism to the observed {111} orientation from an initial {110} nucleation plane. All these studies of dislocation loop energies and morphologies, using either elasticity theory or computer simulation techniques, contributed to the understanding of the microstructural properties in irradiated metals.

The radiation-produced point defects, in addition to forming new clusters or recombining with one another, can also migrate to existing dislocations in the crystal. The interaction of such point defects with dislocations is thus important and our understanding of this interaction was assisted by the fact that Bullough and Newman(9) had extended the theory in a series of papers prior to 1964, culminating in an extensive review of the subject by Bullough(10) for the 1968 Harwell symposium on 'Point Defect- Dislocation Interactions'. A particularly attractive piece of research in this series of papers was an accurate prediction of the rate of oxygen depletion in neutron-irradiated niobium by Bullough, Tucker and Williams(11). The interstitial oxygen was supposed to migrate under the influence of the dislocation stress fields towards the interstitial dislocation loops formed by the irradiation and was thus a very nice manifestation of the point defect-dislocation drift interaction previously familiar to metallurgists through strain ageing and yield-point drop phenomena in unirradiated metals.

In addition to these studies of the properties of clusters, research into basic damage processes was also undertaken in the Division. In 1966 von Jan(12) extended the earlier pioneering work of Thompson and Nelson at Harwell, by determining the way that of the distribution of point defects remaining after high energy collisions depended upon the interatomic potential. He also extended the theory of channelling by calculating the distribution of ranges of channelled ion beams(13). Cheshire et al(14) further improved our understanding of channelling by successfully explaining, with a semi-classical Firsov model, the oscillatory dependence of electronic stopping cross-sections on the atomic number of the channelled ions. At about the same time Norgett began his studies of collison cascades. The theory of recoil damage in solids had begun at Harwell with the early work of Kinchin and Pease(15) and Norgett set out to improve immobile and bulk recombination resulted in a large loss of the primary defects. While at higher temperatures the equilibrium thermal vacancy concentration could exceed the steady state radiation-produced vacancy concentration, in which case any voids would be unstable and would shrink by emitting vacancies.

The huge growth of observational data on void swelling that now occurred, of course, provided many interpretative problems for the theoreticians both at Harwell and elsewhere. Amongst these observations, and a particularly beautiful one, was the observation by Evans in 1970(20) that the voids formed in molybdenum under nitrogen ion irradiation could adopt a perfect body-centred superlattice configuration which was aligned parallel to the underlying atomic lattice. Such void lattices, always parallel to the host atomic lattice and with the same centering as the host lattice, have now been observed in a large number of f.c.c. and b.c.c. irradiated metals. The theory of this void lattice formation has been developed entirely at Harwell(21). The essential reason for the formation of void lattices is that the voids have a weak elastic interaction with each other which is sufficient to encourage the nucleation and growth of ordered regions of voids during the irradiation. Furthermore, the coincidence with the underlying atomic lattice arises because the voids are invariably faceted and thus convey the microscopic symmetry as well as the long range (cubic) symmetry of the atomic lattice to each other via the host medium.

The present rather sophisticated theory of void swelling, which includes the evolution of the microstructure, began to develop in 1972 with a series of papers by Brailsford and Bullough(22), who took advantage of the relative computational simplicity of the quasi-chemical rate theory of defect reactions. The strengths of the various sink types that together define the microstructure were calculated by embedding them in an effective 'lossy' continuum by a method that is a generalization of Maxwell's construction of an effective homogeneous dielectric medium for the representation of an inhomogeneous dielectric. At the time of the first observation of voids in reactor components it was realised by Nelson and his colleagues$^{(23)}$ at Harwell that accelerated damage rates provided by electrons in the H.V.E.M. or by heavy ions from the VEC could provide an important means of studying the swelling phenomenon. However, to correlate the swelling under these simulation conditions with the swelling to be expected under actual fast-neutron irradiation requires a detailed understanding of the sensitivity of the microstructure to the irradiation environment. For example, it is necessary to perform the higher dose-rate experiments at a higher temperature than the neutron damage experiments and the effect of differences in the recoil spectra must be understood. This latter point was discussed theoretically in 1975 by Bullough, Eyre and Krishan$^{(24)}$ who included in the general rate theory the possibility of vacancy loop formation from the depleted zones within the cascades. This work thus enabled a quantitative correlation to be made between the simulation data and appropriate neutron data. Many other features are now included in the current rate-theory model of swelling. These include the effects of high rates of generation of He gas atoms, particularly at high temperatures, which are particularly relevant in studies of the 'first wall' of a fusion reactor where the generation rate of helium by n/α reactions from 14 MeV neutrons will be very much higher than in the fast reactor core components$^{(25)}$. Detailed studies of void nucleation and void size distributions have also been made(26). All this work now provides a comprehensive rate theory of swelling that can include all conceivable microstructural effects. This theory has been incorporated in actual component design codes(27).

In addition to its importance for void swelling, the interaction between point defects and dislocations is also important in irradiation creep. The necessity for at least two different types of sink if a net flow of point defects to one sink is to occur under irradiation has been recognised for the void growth problem. Similarly for irradiation creep the applied stress can induce differing interactions between the dislocations and the point defects, depending on the relative orientation of the dislocations to the applied stress direction. This is the basis of the so-called SIPA irradiation creep (stress-induced preferred absorption) theory(28) which yields quantitative agreement with much of the relevant creep data. Again the sensitivity of this creep to impurity trapping and other microstructural features has been discussed recently by Bullough and Hayns(29). Finally the rate-theory approach has recently been adapted by Bullough and Wood(30) and by Buckley, Bullough and Hayns(31) to interpret the growth of the hexagonal metals zirconium and Zircaloy. Again the importance of impurity trapping is clear in these materials; we believe the substitutional tin atoms in Zircaloy trap vacancies and are thereby largely responsible for the resistance of Zircaloy to radiation growth compared to pure zirconium.

In conclusion, the last 25 years have seen a continual and significant contribution from members of Theoretical Physics Division to the theory of radiation damage in metals. In terms of numbers of researchers our effort has always been small, but nevertheless we believe our impact on the fundamental understanding of radiation damage will continue to be important. We look forward particularly to a growth in effort on damage studies for the fusion reactor environment where new and exciting problems are certain to appear.

1. C. Cawthorne and E.J. Fulton, Nature 216, 575 (1967).

2. G.W. Greenwood, A.J.E. Foreman and D.E. Rimmer, J. Nucl. Mat. 4, 305 (1959).

3. S.N. Buckley, "Properties of Reactor Materials and the Effects of Radiation Damage", 1962, Ed. D.J. Littler (Butterworths, London) p.413.

4. S.N. Buckley, A.E.R.E. Report R-5262 (1966).

5. N. Meissner, Ph.D. Thesis, University of Surrey (1973).

6. R. Bullough and A.J.E. Foreman, Phil. Mag. 9, 315 (1964).

7. B.L. Eyre and R. Bullough, Phil. Mag. 12, 31 (1965).

8. R. Bullough and R.C. Perrin, Proc. Roy. Soc. A305, 541 (1968).

9. R. Bullough and R.C. Newman, Proc. Roy. Soc. A249, 427 (1959); Phil. Mag. 6, 407 (1961); Proc. Roy. Soc. A266, 198 (1962); Proc. Roy. Soc. A266, 209 (1962).

10. See also: R. Bullough and R.C. Newman, Rep. Prog. Phys., 33, 101 (1970).

11. R. Bullough, J.T. Stanley and R.D. Williams, Met. Science Jnl. 2, 93 (1968).

12. R. von Jan, A.E.R.E. Report R-5269 (1966).

13. R. von Jan, Phys. Rev. Letters 18, 303 (1967).

14. I. Cheshire, G. Dearnaley and J.M. Poate, Proc. Roy. Soc. A311, 47 (1969).

15. G.H. Kinchin and R.S. Pease, Rep. on Prog. in Physics 18, 1 (1955).

16. M.J. Norgett, M.T. Robinson and I.M. Torrens, A.E.R.E. Report TP-494 (1972).

17. R. Bullough, D.M. Maher and R.C. Perrin, Nature 224, 364 (1969).

18. R. Bullough, B.L. Eyre and R.C. Perrin, Jnl. of Nucl. Appl. and Tech., 9, 346 (1970).

19. R. Bullough and R.C. Perrin, Proc. of B.N.E.S. European Meeting on "Voids Formed by Irradiation of Reactor Materials", Reading, March 1971 (Ed. S.F. Pugh, M.H. Loretto and D.I.R. Norris).

20. J.H. Evans, Nature 229, 403 (1971).

21. K. Malén and R. Bullough in ref. 19, p.109; A.M. Stoneham, J. Phys. F. (Metal Physics) 1, 778 (1972); V.K. Tewary and R. Bullough, J. Phys. F. (Metal Physics) 2, L69 (1972).

22. A.D. Brailsford and R. Bullough, J. Nucl. Mat., 44, 121 (1972); Phil. Mag. 27, 49 (1973); J. Nucl. Mat. 48, 87 (1973); Nucl. Met. 18, 493 (1973).

23. R.S. Nelson, D.J. Mazey and J.A. Hudson, J. Nucl. Mat. 37, 1 (1970).

24. R. Bullough, B.L. Eyre and K. Krishan, Proc. Roy. Soc. A346, 81 (1975).

25. R. Bullough, M.R. Hayns and M.H. Wood, J. Nucl. Mat. 85 and 86, 559 (1979).

26. M.R. Hayns and M.H. Wood, A.E.R.E. Report TP-789 (1979).

27. M.R. Hayns and R. Bullough, Proc. of 9th Int. Symposium on "Effects of Radiation on Structural Materials", ASTM: STP.683, 1978, p.143.

28. R. Bullough and J.R. Willis, Phil. Mag. 31, 855 (1975); R. Bullough and M.R. Hayns, J. Nucl. Mat. 57, 3 (1975).

29. R. Bullough and M.R. Hayns, TRANSAO 27, 1 (1977).

30. R. Bullough and M.H. Wood, Proc. of Int. Conf. on "Fundamental Mechanisms of Irradiation Induced Creep and Growth", Chalk River, Canada, 8-10 May 1979; J. Nucl. Mat. to appear.

31. S.N. Buckley, R. Bullough and M.R. Hayns, A.E.R.E. Report R-9565 (1980); J. Nucl. Mat., to appear.

Chapter XVI

RADIATION DAMAGE IN NON-METALS

A. M. Stoneham

1. Introduction: Early History

Radiation effects in non-metals gave some of the earliest evidence for radioactivity, for Bequerel's discoveries in 1896 were based on effects seen in silver halide emulsions. Interpretations of the damage on an atomic scale came only much later. It was commonly held in the nineteenth century (e.g. Ruskin(1) 1865) that crystal defects could only be chemical impurities or faults in crystal habit. The ideas of intrinsic defects, like vacancies, interstitials, or dislocations, were far less natural than they appear now.

Radiation damage involves several distinct aspects. One is defect production: the various mechanisms of energy loss and the way in which some of this energy may be used to create defects. Another aspect determines which primary defects emerge and the ways in which defects annihilate or larger defect aggregates grow. A third concerns the way the defects affect other observable properties. This determines both how one monitors and studies damage and the consequences of damage for any practical application. The range of phenomena is extremely wide, and the short and incomplete list in Table I merely samples this diversity.

Since this short article is part of a volume to mark the twenty-fifth anniversary of the U.K.A.E.A., I have chosen to emphasise what has been done at Harwell, and stressed the contribution of the Theoretical Physics Division. Obviously, many of the early advances were led by experiment, as in the work on graphite and UO_2. Equally, some of the important aspects of theory came from other Divisions: the early ideas of radiolysis from Varley, the group theory of defects in applied fields from Runciman and Hughes, and Pooley's excitonic model of colour centre production. A personal view like the present note can never be encylopaedic, though it can isolate some of the themes which have emerged.

Table 1

Radiation Damage in Non-Metals

System	Example of Phenomena Affected by Radiation
Alkali halides and Alkaline Earth Fluorides	Fundamental studies; dosimetry; waste storage; fluorescent screens; information storage
Oxides, e.g. MgO, TiO_2, UO_2	Reactor fuel studies; corrosion studies
More complex oxides	Glasses for waste disposal; transducers for non-destructive testing
Semiconductors	Radiation detectors; device components; fundamental studies; solar cell behaviour
Graphite	Reactor moderator behaviour
Polymers and other organics	Insulators in a radiation environment

2. Mechanisms of Damage

Many important ideas were established by the mid-1950's. Those surveyed by Kinchin and Pease[2] in 1955 still underly much of what is accepted in the 1970's. One such basic component is the idea of a displacement energy, the minimum kinetic energy one must supply to a lattice atom to cause displacement. Another is the idea of recombination, in which a displaced atom may move rapidly to a vacancy, possibly the one at its own original site. There are many obvious ramifications too. One must consider the energy transfer from a fast particle to a stationary one. If it is less than threshold, heat is generated and the fast particle slowed. There are many more complicated models, including ones which allow anisotropic displacement energies, different thresholds for each sublattice, indirect mechanisms, and so on. The principal anisotropic effects, however, are quite different. Both were anticipated theoretically, 'focussed collision sequences' by Silsbee(3) and 'channelling' by Robinson(4).

The idea of focussed collision sequences is this. If an atom in a close-packed row is struck in a direction along the row, or lying close to it, there can be a series of momentum transfers which leave a vacancy at the site of the struck atom and a distant interstitial. Any one atom moves only a short distance, yet separated defects result. The large separation reduces the chance of direct recombination of interstitial and vacancy. The idea of channelling also involves collimation. Here, however, a fast particle moves along empty channels, guided by the rows of adjacent host atoms. This idea has been exploited in Harwell's extensive work on ion implantation of semiconductors and metals. Other articles in this volume discuss related topics, notably the ones by John Briggs on interatomic collisions, by Ron Bullough on damage in metals, and by Michael Norgett on modelling.

One of the most important ideas confined to non-metals has been that of photochemical damage. In many alkali halides one can form vacancies and interstitials by band-gap optical excitation. The key to this process is the non-radiative decay of the exciton, combined with a focussed collision sequence to separate the components. This was first realised in 1966 by Pooley[5], who gave a detailed semi-quantitative model of the phenomenon. More recently, the theory of the whole process has been re-analysed in a way which includes predictions of the energy surfaces and of the states involved (Itoh, Stoneham and Harker[6]). This is a good example of theory resolving issues inaccessible by experiment alone, for the damage process is very rapid and involves many different electronic states. It is also a good example of work with wider relevance, for the process has many parallels with the ionisation-enhanced and recombination-enhanced phenomena observed in semiconductors.

The successful quantitative analysis of the complicated processes of radiolysis points to one major change in the last ten years, namely the computer revolution. In the mid 1960's it was a rare calculation which handled two electrons on two centres self-consistently. By the mid 1970's such a calculation was routine. One could concentrate on the physics in the reasonable certainty that the mechanical part of the calculation could be solved. Partly this came from hardware developments, though much came from new computer codes. The calculation of F-centre production illustrates the importance of software. It used the MOSES code to treat one- and two-electron excited states of 57 ions self-consistently, these ions being embedded in an array of point-ions; the ion positions were themselves obtained from separate HADES calculations, and the results were cross-checked with the PRISM and ATMOL codes. Michael Norgett discusses the various major codes elsewhere; suffice it to say that they have made it possible to cope with the very varied demands of the wide range of work at Harwell. One can make a serious attempt to look at the problems of real physical interest.

3. Theory of Defects in Solids

By 1940, when Mott and Gurney's seminal volume[(7)] appeared, experiment and theory could already give a convincing picture of some of the key defects in ionic crystals. Thus Seitz' famous review of radiation damage in ionic crystals[(8)], published in the same year as the Authority was formed, dealt with a quite mature field. The situation was very different for semiconductors. The major developments in the effective-mass theory of electrically-active shallow impurities had just been made[(9)]. The key experiments on vacancies and interstitials in these materials - notably spin resonance work on silicon and optical studies of diamond - were still to come, and very little of the early phenomenology has survived later developments. Table 2 shows only a sample of the important defects; many of these documented defects have been studied in the Theoretical Physics Division[(10)].

There is one important division of defects. Some, like the F-centre in an alkali halide (an anion vacancy plus a trapped electron), involve defect electrons, localised at the defect, whose properties must be calculated from the Schrödinger equation, or some approximation to it. Others, like the simple anion vacancy in an alkali halide, involve only closed-shell ions, without defect electrons. Their properties can be obtained much more simply within the shell model. This model is combined with empirical interatomic potentials and polarisabilities and avoids explicit solution of the Schrödinger equation.

Table 2
Systematics of Defects in Solids

This table compares defects by type in the best-known crystalline solids. It oversimplifies in several respects, notably by ignoring alternative crystal structures. Charge states of vacancies are defined by V^{N+} meaning an ion of charge (N-) has been removed; for interstitials, I^{M+} means an ion of charge (M+) is inserted. Traditional labels are given when appropriate.

Type of Defect	Ionic			Covalent	
	Alkali Halides	Oxides	Other II-VI	III-V	Group IV
ANION VACANCY	$V^- = F'$ $V^o = F$ $V^+ =$ = simple vacancy	V^- reported $V^o = F'$ $V^+ = F^+$ V^{++}= simple vacancy	F',F^+ and "simple vacancy" reported sometimes with impurities	Mainly associates or unconfirmed reports	Diamond (V^+), V^o = GRI,V^- = ND1 Silicon (V^{2+}),V^+,V^o,V^-,V^{2-} Germanium V^o,V^-, not V^+(?)
CATION VACANCY	$V^o = V_F$ V^- = simple vacancy	V^o V^- V^{2-}= simple vacancy	V^o V^- (V^{2-} = simple vacancy)	V^{2-} claimed in GaP	

POLYVACANCIES	Divacancy;
	M,R,N centres;
	Colloids
		Anion vacancy aggregates; voids		Colloids?	Divacancies in various charge states; Chains of vacancies;
					Voids

INTERSTITIALS	I^o = H I^+ in fluorites	I^{2+}in fluorites	Evidence very indirect, though dislocation loops are easily found		Inferred from reactions under irradiation and from "swirl" defects at high temperatures
OTHER INTRINSIC DEFECTS	Halogen molecule in anion-cation divacancy; Self-trapped hole; Self-trapped exciton	Vacancy-interstitial complexes; Shear planes Small polarons in some systems		Antisite defects	Atoms with "wrong" co-ordination in amorphous materials.
IMPORTANT IMPURITIES	Molecular ions: OH^-,O_2^-; Amphoteric: Tl, halogen, alkali Hydrogen (many forms)	Alkalis (acceptors) and Halogens (donors) Transition metal ions (many charge states)		Shallow donors, acceptors and isovalent (e.g. N in GaP) impurities; Deep centres (transition metals, Au, etc); "Benign" impurities like H or O	

This second class of defect is often the most important. The properties of the principal defects produced by radiation damage of simple ionic crystals, or of fuels like UO_2, can be found by relatively simple calculations. The key codes in this area have been the HADES and PLUTO codes, and these have been developed and used imaginatively and extensively by Norgett and Catlow and their co-workers. The results give enthalpies and entropies of formation and diffusion, energies of small polaron motion, charge transfer energies, and so on. From the results follow the relative stabilities of defects and the critical energies in different processes. Some of the most important conclusions concern electron and ionic transport in transition metal and actinide oxides. These illustrate two aspects of the way theory complements experiment. Firstly, the quantitative predictions allow one to distinguish between intrinsic and extrinsic effects in what may be unavoidably 'dirty', i.e. impure, crystals. Secondly, the theory can probe regimes which are inconvenient experimentally, like very high temperatures, or very short times.

When solution of the Schrödinger equation is unavoidable, it can be made relatively painless by use of suitable codes. Saunders' ATMOL, and Harker's PRISM, SEMELE and MOSES have all been powerful in resolving questions which experiment alone finds taxing. Examples here include the structure of the self-trapped exciton and its relation to radiation damage(6), the mechanisms of ionisation-enhanced motion of interstitials in diamond(11) (and, by implication, in silicon) and the solvation of chemical impurities in liquid sodium(12).

The comments on computer methods should not imply that all theory is computation. The computer is merely a tool, though, as a tool, it has many advantages, since one can make intuitive ideas quantitative. One of the roles of theory is to provide a framework in which results can be understood, and this conceptual framework is quite distinct from computation. Analytical theory too has remained very productive. Elasticity theory is a traditional, yet fertile, area. Alan Lidiard and Uma Jain[13] have proved the value of the rate theory of the growth of defect aggregates (see Ron Bullough's paper) in non-metals in their analysis of Hobbs' data on heavily-irradiated ionic materials[14]. And the theory of non-radiative transitions in all its aspects has always been primarily analytical[15].

4. Summary

In the years between 1954 and the end of the 1960's the main thrust in the radiation damage of non-metals was model-building: trying to devise defect models and mechanisms that were qualitatively acceptable, compiling systematic data, and feeling grateful for any quantitative success. The early 1970's made greater quantitative demands. Computer techniques made theory more powerful, so that it became a more useful (and often equal) partner with experiment. In many cases one could predict defect properties accurately, so that one could distinguish between different defect models which were hard to tell apart by experiment alone. In the late 1970's, the most important aspect has changed again. Now it has moved towards mechanisms of defect processes, especially in cases where experiment by itself is limited by timescale, by complexity, by the unintentional impurities inevitable in real crystals, or by the extreme conditions required. There can be little doubt that such developments will remain of importance for some time to come.

1. J. Ruskin, "The Ethics of the Dust" (George Allan, 1865). See especially Section 47, where the two important passages are these: "I shall have to tell you of the faults of the crystals ... And some have a great many faults; and some are very naughty crystals indeed." "... but crystals have a limited, though a stern, code of morals; and their essential virtues are but two; - the first is to be pure, and the second is to be well-shaped."

2. G.H. Kinchin and R.S. Pease, Rep. Prog. Phys. 18, 1 (1955).

3. R.H. Silsbee, J. Appl. Phys. 28, 1246 (1957).

4. M.T. Robinson and O.S. Oen, Appl. Phys. Lett. 2, 30 (1963).

5. D. Pooley, Proc. Phys. Soc. 87, 257 (1966). Observations of photochemical damage include work as early as that of A. Simakula, Z. Phys. 63, 762 (1930).

6. N. Itoh, A.M. Stoneham and A.H. Harker, J. Phys. C.10, 4197 (1977).

7. N.F. Mott and R.W. Gurrey, 1940, "Electronic Processes in Ionic Crystals" (Oxford University Press).

8. F. Seitz, Rev. Mod. Phys. 26, 7 (1954).

9. The earliest contributions appear to be in ref. 7 and in H.A. Bethe, M.I.T. Radiation Laboratory Report 43-12 (1942), cited by S.T. Pantelides, Rev. Mod. Phys. 50, 797 (1979). The two key original papers are C. Kittel and Mitchell, Phys. Rev. 96, 1488 (1954) and W. Kohn and J.M. Luttinger, Phys. Rev. 97, 883 (1955); Kohn's 1957 review (Sol. St. Phys. 5, 257) has been of especial importance.

10. See, for example, A.M. Stoneham, 1975, "Theory of Defects in Solids" (Oxford University Press) and A.M. Stoneham, Adv. Phys. 28, 457 (1979).

11. A. Mainwood, A.M. Stoneham and F.P. Larkins, Sol. St. Electr. 21, 1431 (1978).

12. A. Mainwood and A.M. Stoneham, Phil. Mag. 37B, 255 and 263 (1978).

13. A.B. Lidiard and U. Jain, Phil. Mag. 35, 245 (1977).

14. L.W. Hobbs, Surface and Defect Properties of Solids, (Chemical Society Specialist Peiiodical Reports) 4, 152 (1975).

15. See, e.g., ref. 10 and A.M. Stoneham, Phil. Mag. 36, 983 (1977).

Chapter XVII

DEFECT MODELLING

M. J. Norgett

1. Introduction

As Alan Lidiard's article in this volume points out, the study of imperfections in solids has developed in the last 25 years to form a distinct branch of solid-state science. The theories of various important phenomena, such as diffusion, radiation damage, mechanical properties, the solid-state chemistry of oxides and corrosion, all depend on the structure of lattice defects or their energies, mobilities and interactions. Until recently, our knowledge of such basic properties was limited, and often based on very simple and intuitive models. However, in the last decade, it has been possible to make extensive and detailed calculations whenever there exists a suitable potential function for the solid; such defect modelling has made substantial contributions to our understanding. These fruitful - if often frustrating - studies have lately taxed the imagination, ingenuity and initiative of various members of Theoretical Physics Division. We have also enticed several of our visitors into this particular web; their different interests and enthusiasms have broadened our outlook and advance.

Any comprehensive review of defect modelling needs to consider various types of calculation. We shall, however, ignore simulations of displacement processes and atomic collision sequences. Ron Bullough in his article shows how such studies predicted and explained important radiation effects such as channelling and focussing. Moreover, we shall also neglect dynamic studies of defect crystals, which have only recently begun to provide new insights. In this review, we consider only calculations of the relaxation about lattice defects; the Division has certainly worked hardest in this area and so far has made most progress here. We will emphasize developments in the techniques required for such calculations because both Ron Bullough and Marshall Stoneham give other examples of how such modelling has increased our understanding of defects and damage processes.

2. Defect Modelling - Principles

The basic principle of a calculation of lattice distortion about a defect is very simple. A region of stable, ideal crystal is first represented by an assembly of atoms with interactions described by a suitable interatomic potential. The perfect crystal is then disturbed by introducing lattice defects so that forces act on the adjacent atoms. Those atoms in a limited region surrounding the defect are relaxed explicitly to equilibrium; the displacements of more distant atoms are neglected; or, much better, are calculated assuming that the material responds as a continuum.

Such simulations thus predict the equilibrium configuration of the atoms in the imperfect crystal. For dislocations, such distortion fields can sometimes be inferred directly from observations in the electron microscope; and the configuration of such extended defects also correlates with mechanical properties of materials. The lattice distortion about point defects is less accessible to direct study, but calculated energies of defect formation and migration do match results of measurements of diffusion or ionic conductivity (see chapter XIV by Alan Lidiard).

Thus far, we have described the basic object of defect simulation and the results to be obtained from these studies. We shall go on to consider specific recent achievements; but first, to give some guide to this history, it is worthwhile to outline general problems and areas where we might look for particular progress.

To begin, we emphasize that any simulation can be no better than the chosen potential model. Interatomic interactions are, of course, commonly deduced from varied data, or obtained using several methods of calculation; but models for defect studies must satisfy particularly stringent criteria. Potentials must describe distorted lattices and be applicable over a range of interatomic separations; they must also be independent of lattice configuration. Defect modelling is thus not only an important application of interatomic potentials; it is also a critical test of models based on other data.

It is therefore exceptional to complete any study with the original model intact. Results of a preliminary defect calculation usually suggest a refinement or recasting of the potentials. This cycle must often be repeated through several interactions; and if this series stubbornly refuses to converge, that particular problem is best set discreetly aside. It is wrong, however, to despair too soon: a lively mind can find endless ways of refining potentials. We might particularly recommend the study of the interaction between a specific ion pair using data for different chemical substances: for example, data for the four rock-salt crystals LiF, NaF, KF and RbF, the fluorites CaF_2, SrF_2 and BaF_2, and the rutiles MnF_2 and MgF_2 together provide complementary information about the F^--F^- potential at a range of interatomic separations and in different crystalline environments. This chemical approach is particularly congenial to the various renegade chemists who have made significant contributions in this field; it is also an illuminating and productive approach. Of course, the study of complex crystal structures has particular difficulties, but we shall see later how these can be overcome.

If we have achieved a satisfactory potential, the calculation of lattice relaxation can begin in earnest. A little experience reveals that this will impose a substantial burden of computation except in studies restricted to the simplest models, and with relaxation of only the immediate neighbours of the defect. If the calculation is not carried out with some finesse, anything more ambitious will stretch the capabilities of even modern computers. We can identify two important opportunities for accelerating such computations that can achieve substantial economies.

First, recall that the crystal is relaxed only in a limited region which is surrounded by a boundary beyond which the distortion is represented by a continuum displacement field. Any improvement in this boundary field allows a reduction in the size of the explicitly-relaxed lattice region, and hence a decrease in computation without loss of precision.

A second opportunity for economy is to develop and apply improved numerical methods. It was our good fortune at Harwell to collaborate closely with the Numerical Analysis Group, particularly in past days when they were closetted with Theoretical Physics Division. Roger Fletcher and Mike Powell record in this volume those advances in the theory of optimization to which they made particular contributions; much of this work had a direct impact on the development of defect modelling.

By employing such methods, it became possible in favourable cases to make essentially complete defect calculations which were not compromised by the limited size of the relaxed region. Such results could therefore be correlated absolutely with characteristics of specific potentials; this allowed a true appreciation of the benefit of improved models. It also became possible to contemplate the study of complex defects; but this was at first a daunting prospect. The tedium of such complex but routine calculations would have blunted the enthusiasm of the most ardent research student. Fortunately, the detailed manipulations in such calculations were subsumed into package programs that could be applied generally to complex defect studies.

Of several such programs developed at Harwell, HADES*, DEVIL** and PLUTO*** have seen most service in defect simulations. HADES brings together a substantial experience of studying defects in ionic crystals. DEVIL is less self-contained but incorporates several flexible methods for exploring extended defects in materials represented by short-range potentials. PLUTO was developed for analysing potential models suitable for complex crystal structures: but has also been used to study very imperfect materials with defect superlattices.

* Harwell Automatic Defect Examination System
** Defect Evaluation in Lattices
*** Perfect Lattice Unrestricted Testing Operation

Such programs have a significance beyond the immediate opportunity provided by improved methods. It becomes possible to carry out a wide variety of calculations with strictly equivalent assumptions; this separates significant and incidental differences in results. Moreover, a single source code can be tested extensively in varied applications to eliminate errors that are difficult to exclude from more restricted programs. Finally, such codes are a tool for general use; work carried on independently outside the Division is the true measure of the value of such programs.

The development of such codes, with particular attention to facilities for simple, flexible data input, of course requires a substantial additional effort. We have therefore generally insisted on charging for copies of such programs. This policy has not always been understood or well received by those who consider a free exchange of computer codes essential for full publication of basic scientific research. By publishing principles of programs, but not explicit details, we have tried to satisfy two conflicting boundary conditions, maintaining scientific respectability without forgoing a valid opportunity for securing very necessary outside income.

3. Historical Preliminaries

Thus far, I have tried to survey defect modelling in general terms, to point out broad opportunities and trends. We can now appreciate the significance of what has been achieved in recent years. The work recorded in this brief review is necessarily selected to illustrate particular themes. In the later period, it draws principally on work that depends on developments at Harwell such as particular computer programs. A brief summary obviously must omit many important achievements, but because in this field the Division has been at the centre of events, we may fairly claim that this review neglects no major area of importance.

The opportunity for defect modelling was realized some ten years after Frenkel and Schottky had implicated lattice defects in matter transport in solids. In 1938, Mott and Littleton(1) published the first calculations of the formation and activation energies of vacancies in alkali halides. These studies were necessarily limited in scope, but they identified precisely the path to further progress and, in particular, the importance of using a proper dielectric boundary region about a charged defect.

In the 1950's and early 1960's the same simple Born model, with polarizable ions, was used to explore various simple defects in ionic crystals. Tharmalingam and Lidiard(2), for example, calculated vacancy-pair formation energies. This work was, we admit, carried out during Alan Lidiard's banishment to Reading; but it marked out a future interest for the Division when he returned.

At much this time, R.A. Johnson(3) in the U.S.A. was bold enough to use simple short-range pair potentials for a study of interstitials in copper and iron. He was rewarded with sensible predictions of the defect configuration, although such models might seem wholly inappropriate as a description of free-electron metals.

4. Particular Calculations: Mainly Metals

By the late 1960's, these foundations could be developed rapidly using the improved computers then available. In 1968, for example, Bullough and Perrin(4) linked the properties of separate point defects with the development of dislocation loops in iron. They simulated clusters of numerous interstitials and recorded the transformation of such aggregates into the nucleus of a perfect dislocation loop with different orientation.

Such methods offered a new opportunity for exploring defects in detail at an atomic level; these techniques were therefore soon used to probe various dislocations; and, in this quest, the Division maintained a prominent place. Perrin, Englert and Bullough(5) modelled dislocations in copper, particularly the glissile edge dislocation that bestows high ductility. Norgett, Perrin and Savino(6) subsequently calculated the separation of the partial dislocations in this configuration for comparison with direct observations in the electron microscope. This was the first appearance of the DEVIL program and of Fletcher's effective conjugate gradient method(7) applied to defect simulation. At this time, as always, we were sustained by the charm, enthusiasm and exuberance of our visitors.

This work at Harwell was complemented by similar studies elsewhere, to which the Division contributed particular skills on specific programs. Thus Perrin joined with Vitek and Bowen(8) at Oxford to study screw dislocations in body-centred-cubic metals. Crocker's group at Surrey University(9) applied the DEVIL code to simulate twin boundaries in similar materials.

We thus spread our bread broadly on the waters and were ourselves sustained when Sinclair brought to the Division his own experience in dislocation modelling(10), and his particular achievements in applying more sophisticated elastic boundary regions in lattice simulations. These general methods had most obvious impact in Sinclair's own study(11) of brittle crack propagation; where the elastic displacement field must reflect the advance of the growing crack.

Sinclair's later work concentrated on models of diamond because simple pair potentials are an inadequate description of a metal surface. In fact, such models seem inadequate for predicting defect energies and this limitation has restricted defect modelling in metals to studies of lattice configurations. There have, nonetheless, been substantial efforts to develop better models based on pseudo-potential theory, but these have had only limited success. For example, Schober, Taylor, Norgett and Stoneham(12) did calculate the characteristic energies of silver and gold substituted in lithium and sodium; pseudo-potential calculations of pair potentials are satisfactory in this case only because the impurities do not provide a serious perturbation of the electron structure of the alkali metal.

Despite this declining interest, those programs developed for studying metals have not been discarded. A variant of the DEVIL code has recently been applied to simulating thin boundaries in solid nitrogen by Venables and his colleagues(13) at Sussex. Bacon's group at Liverpool(14) has used a similar code to study the crystal structure of polyethylene. Thus a computer program that incorporates flexible general methods can have an expanding range of application and an extended useful life.

5. Particular Calculations - Ionic Crystals and Oxides

We have thus brought this review of the simulation of extended defects in metals, and other materials where short-range potentials are suitable, through to the present day. We have, however, deferred any consideration of comparable recent advances in modelling ionic materials, which warrant a separate discussion. This is because the problems and opportunities are rather different; thus the long-range Coulomb potential is simple in form but more difficult to compute; any suitable model must also describe ionic polarization, but such models are better than those available for other materials and do provide precise defect energies.

The renaissance in this field also began in the late 1960's and depended on the developments in computers at that time. About this date, Norgett and Lidiard(15) used a simple Born model to calculate activation and trapping energies of inert gases in alkali halides; such results provided a clear interpretation of the observed diffusion. In the U.S.A., a group at Brookhaven(16) simulated the observed off-centre configuration of Li^+ substituted in KCl. However, it seemed that such simple models, based on polarizable point ions, were not wholly appropriate.

Both studies were very time-consuming and it was clear that better numerical methods would be necessary if there was to be much further progress. Norgett and Fletcher(17) soon demonstrated the particular efficiency of the variable metric methods for calculating lattice relaxation in ionic materials. In this way, computation times could be reduced by factors of up to one hundred, and it became possible to contemplate calculations for ionic crystals containing complex defects.

It was necessary first, however, to dispose of problems with the potentials. A little before this time, Boswarva and Lidiard(18) had calculated Schottky energies for alkali halides using various models; they had achieved results in generally good agreement with experiment, broadly independent of their detailed assumptions. However, they had relaxed only immediate neighbours of the defect; when Scholz(19) considered a larger region of lattice, the answers were much worse. With new opportunities to carry out a variety of such extended calculations, it became clear that this problem was characteristic of the polarizable point-ion model then in use.

The remedy followed immediately upon the recognition that, in the point-ion model, the ionic polarization is not properly damped as ions are displaced and overlap. The model lattice is thus too polarizable, so that the explicit region of crystal does not match the surrounding boundary - the continuum, of course, has a proper dielectric response calculated according to the method established by Mott and Littleton in 1938. With a point-ion model, the calculated Schottky energies therefore decrease as the region of explicit lattice is enlarged. To avoid this, Faux and Lidiard(20) substituted a shell model, with consistent dielectric behaviour; they thus obtained accurate calculated Schottky energies essentially independent of the size of relaxed region.

The powerful numerical methods and the shell model were the basic foundations of the HADES program. Catlow and Norgett(21) first applied this code in a shell-model study of simple point defects in CaF_2. Catlow(22) then employed the program's full capabilities on a more demanding problem; he demonstrated the stability of the complex clusters of vacancies, interstitials and Y^{3+} ions that had been postulated to explain the structure of CaF_2 containing substantial amounts of YF_3.

The subsequent extension of such calculations to oxides was a particular achievement of the shell model; previous polarizable point-ion calculations for such materials had been very unsatisfactory. Catlow and Lidiard(23) extended Catlow's work on CaF_2 to flourite oxidees and thus clarified the thermodynamics of non-stoichiometric UO_2. A preliminary study of MgO(24) provided a pathway to calculations for MnO, FeO, CoO and NiO which have properties important in determining corrosion. These latter studies(25) have been carried out in co-operation with I.C.I.'s Corporate Laboratory, where Mackrodt's group are concerned with semi-conducting oxides that are often active in catalysis.

These studies of transition metal oxides have in fact considered mass transport and also electronic conductivity, which is determined by the hopping of electron holes localized as small polarons to form distinct M^{3+} ions. The temperature dependence of the polaron hopping depends on lattice energies that can be computed using the HADES program.

In the last few years, the areas of application of such modelling have multiplied and we can consider only some of the more fruitful branches of this tree. Although the alkali halides have been much studied, they remain important for developing new methods and models and because it is possible to make detailed comparisons with experiment. Catlow, Diller and Norgett(26) have based new potentials for these substances on a broad and consistent range of data; these models have been used in calculations of intrinsic defect energies(27), the behaviour of substitutional ions(28), and studies of irradiation-induced defects(29) which cluster and form dislocations. We have sought expert help - and made new friends - in preparing detailed comparisons of such results with experiment. Corish and Jacobs have contributed their detailed experience to our analysis of conductivity and diffusion in alkali halides which is based on defect calculations. Hobbs has helped correlate the results for irradiation-induced defects with his own studies of halides in the electron microscope.

For oxides, the present interest is in complex defect clusters and defects in more complicated crystal structures. Catlow and Fender(30) explored the specific stability of clusters in FeO that correspond to the basic units of the spinel structure. Gourdin and Kingery(31) made similar studies of Al^{3+} and Fe^{3+} in MgO and have used calculated binding energies to make sense of a complex defect chemistry. In such ceramic materials, it is much more difficult to identify particular defects experimentally, and the value of simulations is enhanced.

The range of different crystal structures that have been explored grows apace. Kilner and Brook, working with Norgett at Harwell(32), have used the HADES program to screen various perovskite oxides in a search for fast ion conductors. James(33) has studied defects in Al_2O_3 and TiO_2. His study with Catlow of highly defective TiO_2(34), where gross departures from stoichiometry are incorporated as ordered lattice shear planes, is one of the most ambitious applications of the simple shell model to defect modelling. This simple approach is still yielding original and exciting results.

Such studies of extended defects in ionic crystals are certainly demanding but at present seem very worthwhile. In this area, Norgett and

Puls(35) simulated edge dislocations in MgO; and then, with Woo(36), calculated binding energies of vacancies at various lattice sites near the dislocation core. On another track, Stewart and Mackrodt(37) adapted HADES to make calculations of surface defects and Tasker(38) at Harwell has recently been developing a more systematic code for such problems. Clearly, this is an area where we may expect energetic activity in the immediate future.

Meanwhile, many interesting problems are still well within the capability of existing codes. The further application of the HADES and PLUTO programs is assured by their transfer to the Science Research Council, whose Daresbury Laboratory now acts as their custodian. The programs are freely available to university groups in the U.K., but others must still accept that the labourer is worthy of his hire.

At Harwell, several of the old guard have passed to nobler - or, at least, new - preoccupations. They have left a legacy to a younger generation eager for emulation. When they come to celebrate the Authority's half centenary, we may have some confidence that there will still be something interesting for them to say about their own achievements in defect modelling.

1. N.F. Mott and M.J. Littleton, Trans. Faraday Soc., 34, 485 (1938).

2. K. Tharmalingam and A.B. Lidiard, Phil. Mag., 6, 1157 (1961).

3. R.A. Johnson, Phys. Rev., 127, 446 (1962) and Phys. Rev., 134A, 1329 (1964).

4. R. Bullough and R.C. Perrin, Proc. Roy. Soc. A, 305, 541 (1968).

5. R.C. Perrin, A. Englert and R. Bullough, in "Interatomic Potentials and Simulation of Lattice Defects", eds. P.C. Gehlen, J.R. Beeler and R.I. Jaffee (Plenum Press, 1972), p.509.

6. M.J. Norgett, R.C. Perrin and E.J. Savino, J. Phys. F (Metal Phys.), 2, L73 (1972).

7. R. Fletcher and C.M. Reeves, Comput. J., 7, 149 (1964).

8. V. Vitek, R.C. Perrin and D.K. Bowen, Phil. Mag., 21, 1049 (1970).

9. P.D. Bristowe, A.G. Crocker and M.J. Norgett, J. Phys. F (Metal Phys.), 4, 1859 (1974).

10. J.E. Sinclair, J. Appl. Phys., 42, 5321 (1971).

11. J.E. Sinclair, Phil. Mag., 31, 647 (1975).

12. H. Schober, R. Taylor, M.J. Norgett and A.M. Stoneham, J. Phys. F (Metal Phys.), 5, 637 (1975).

13. G.J. Tatlock, R. Mevrel and J.A. Venables, Phil. Mag., 35, 641 (1977).

14. N.A. Geary and D.J. Bacon, in "Computer Simulation for Materials Applications", eds. R.J. Arsenault, J.R. Beeler and J.A. Simmons, Nuc. Metall., 20, Pt. I, 479 (1976).

15. M.J. Norgett and A.B. Lidiard, Phil. Mag. 18, 1193 (1968).

16. W.D. Wilson, R.D. Hatcher, G.J. Dienes and R. Smoluchowski, Phys. Rev. 161, 888 (1967).

17. M.J. Norgett and R. Fletcher, J. Phys. C (Solid St. Phys.), 3, L190 (1970).

18. I.M. Boswarva and A.B. Lidiard, Phil. Mag., 16, 805 (1967).

19. A. Scholz, Phys. Stat. Solidi, 25, 285 (1968).

20. I.D. Faux and A.B. Lidiard, Z. Naturforsch., 26a, 62 (1971).

21. C.R.A. Catlow and M.J. Norgett, J. Phys. C (Solid St. Phys.), 6, 1325 (1973).

22. C.R.A. Catlow, J. Phys. C (Solid St. Phys.), 6, L64 (1973).

23. C.R.A. Catlow and A.B. Lidiard, in 'Thermodynamics of Nuclear Materials 1974', Vol. II, I.A.E.A., Vienna, 1975, p.27.

24. C.R.A. Catlow, I.D. Faux and M.J. Norgett, J. Phys. C (Solid St. Phys.), 9, 419 (1976).

25. C.R.A. Catlow, W.C. Mackrodt, M.J. Norgett and A.M. Stoneham, Phil. Mag., 35, 177 (1977) and Phil. Mag., 40, 161 (1979).

26. C.R.A. Catlow, K.M. Diller and M.J. Norgett, J. Phys. C (Solid St. Phys.), 10, 1395 (1977).

27. C.R.A. Catlow, J. Corish, K.M. Diller, P.W.M. Jacobs and M.J. Norgett, J. Phys. C (Solid St. Phys.), 12, 451 (1979).

28. C.R.A. Catlow, K.M. Diller, M.J. Norgett, J. Corish, B.M.C. Parker and P.W.M. Jacobs, Phys. Rev. B, 18, 2739 (1978).

29. C.R.A. Catlow, K.M. Diller and L.W. Hobbs, Phil. Mag., to be published.

30. C.R.A. Catlow and B.E.F. Fender, J. Phys. C (Solid St. Phys.), 8, 3267 (1975).

31. W.H. Gourdin and W.D. Kingery, J. Mater. Sci., 14, 2053 (1979).

32. J.A. Kilner, P. Barrow, R.J. Brook and M.J. Norgett, J. Power Sources, 3, 67 (1978).

33. R. James, Ph.D. Thesis, London, 1978 and A.E.R.E.-Harwell Report TP.814.

34. C.R.A. Catlow and R. James, Nature, 272, 603 (1978).

35. M.P. Puls and M.J. Norgett, J. Appl. Phys., 47, 466 (1976).

36. M.P. Puls, C.H. Woo and M.J. Norgett, Phil. Mag., 36, 1457 (1977).

37. W.C. Mackrodt and R.F. Stewart, J. Phys. C (Solid St. Phys.), 10, 1431 (1977).

38. P.W. Tasker, J. Phys. C (Solid St. Phys.), 12, 4977 (1979).

Chapter XVIII

NUMERICAL ANALYSIS

M.J.D. Powell

One of the most keenly debated questions on research in numerical analysis is whether or not it is sensible to aim the research at particular applications, in order to ensure that the work is useful. I am convinced that it can be disadvantageous for the immediate needs of computer users to have a strong influence on research into general algorithms, and this opinion is mainly due to the successful development of numerical analysis that has taken place at Harwell during the last 25 years. Except for the National Physical Laboratory, there is no government or industrial research laboratory in Britain that has achieved a reputation in the subject that is as high as the one that Harwell has gained. Mainly, this is due to the fact that many of the Harwell algorithms have been applied successfully to a wide range of scientific calculations throughout the world. Also the quality of the research at Harwell is highly regarded by numerical analysts generally.

These achievements have come from Harwell's response to the transformation that has taken place since the 1950's in the use of computers for scientific calculations. The increase in computer power has been enormous, and the number of computer users has multiplied greatly, but the number of numerical analysts employed at Harwell is about the same. The way in which the nature of the work of the Numerical Analysis Group has changed to meet these needs will be reviewed.

When I first joined Harwell in 1959, we used a Ferranti Mercury computer, which was excellent compared with previous equipment. However, the speed of the machine was such that it was necessary to use it efficiently on most calculations, particularly because of the smallness of the core store. Therefore many of the staff of the then Mathematics Group in Theoretical Physics Division at Harwell, including myself, were employed to give direct help to scientists, so our main activities were problem solving and the writing of computer programs. One important exception to helping scientists directly, however, was the provision of sub-programs for the mathematical procedures that occur in many different applications, for example the calculation of the eigenvalues and eigenvectors of a symmetric matrix. We usually turned to the techniques of hand computation and to books for suitable procedures - I found Hildebrand's volume on Numerical Analysis and the N.P.L. publication "Modern Computing Methods" particularly useful. Therefore, much of this work was also computer programming.

We seemed to have sufficient staff until the early 1960's, when about half of the group moved to the newly formed Atlas Computer and Culham Laboratories. Instead of using Mercury, most Harwell computing at that time was being done on I.B.M. machines at Risley and at Aldermaston. Hence the links between numerical analysts and computer users at Harwell became very weak indeed, so the majority of computer users did not expect any expert help with their numerical calculations. In fact, I found myself giving more of my time to the research of American visitors to the Theoretical Physics Division than to the needs of Harwell computer users! Some of the remaining numerical analysts were turning towards computer science. Time has shown that this apparently unfavourable situation was exactly right for the foundation of the highly successful distribution of activity in numerical analysis that is now present at Harwell.

Several benefits came rapidly. Because the scientists at Harwell did not depend on much help from the staff of the Mathematics Group, many of them learnt Fortran and wrote their own computer programs. Thus a practical knowledge of computing spread through the Divisions of A.E.R.E., and early experience was gained of an environment, which is typical today, where the ratio of computer users to numerical analysts is many hundreds to one. Two inefficiencies of the independent computing soon became obvious. First, there was duplication in the writing of computer programs for standard mathematical calculations and, second, inferior numerical methods were often used. Therefore much of the work of the numerical analysts at Harwell was directed towards providing the users with good Fortran subroutines for their mathematical calculations.

Thus the Harwell subroutine library was born in 1963, and it has become the main interface between the work of the Numerical Analysis Group and the contribution that this work makes to scientific research at Harwell. It is now usual for computer users to turn to the Harwell library for the numerical algorithms that they require. Yet the value of the library grew slowly, because most people who have had to develop their own numerical methods are reluctant initially to rely on a general procedure instead.

When the library started we accepted subroutines from many people and, instead of the formal procedures that are current today to achieve quality and accuracy, it was made clear to the contributors that they had the responsibility for the efficacy of their routines. About ninety subroutines were available by the end of 1963 and about one hundred and fifty by the end of 1965. Our early contributors included Ann Bailey (special functions), Alan Curtis (rational approximation), Peter Hallowell (two-point boundary-value problems), Tony Hearn (Gaussian quadrature), Mike Hopper (linear least squares and linear programming), Don McVicar (Runge-Kutta and random number generators), Sid Marlow (special functions), Lewis Morgan (sorting), John Soper (3j- and 6j-coefficients), Ted York (linear equations, eigenvalues and polynomial fitting) and myself (optimization). Due to these contributions and several others, the first library catalogue (AERE Report M-1748) shows that most of the standard methods for numerical calculations were included by the beginning of 1966, so the library was quite suitable for the large number of independent computer users at Harwell.

Before 1966 most of the numerical methods were taken from the published literature, but the Fortran programs were written by the contributors (mostly at Harwell, but also at Culham and the Atlas Computer Laboratory). Some new algorithms, however, were also included. For example, I provided some successful techniques for unconstrained optimization that are mentioned by Roger Fletcher in his article here. Once the library was established it was possible to give even more time to the creation of new numerical methods, so we tried to find out from users what methods would be particularly useful to their work. Because they were reluctant to suggest new fields of research, the areas of study were chosen by the staff of the numerical analysis group. Some of the routines that were added to the library as a resul: of this work are given below. The list covers the period from 1966 until 1973, when the Numerical Analysis Group left Theoretical Physics Division to join the newly formed Computer Science and Systems Division.

In 1966 and 1967, Alan Curtis and I studied the advantages and disadvantages of using cubic splines in approximation calculations, and we provided subroutine TS01A for representing a mathematical function to prescribed accuracy(1), and VC03A for data fitting(2). In 1968 and 1969 I developed an algorithm for solving systems of non-linear equations without calculating derivatives, when a good initial estimate of the solution is not available, namely NS01A(3), and I extended this work to provide subroutine VA05A also, which is an algorithm that is used frequently for non-linear least-squares calculations. Meanwhile, the time that Alan Curtis could spend on numerical analysis was given mostly to the difficult and important problem of choosing step-lengths automatically in the numerical solution of ordinary differential equations, which yielded the successful subroutine DC01A early in 1970. Roger Fletcher had joined us in 1969 and he contributed immediately VE01A, which minimizes a general function subject to linear constraints on the variables. In the following two years he developed a new method for quadratic programming(4), VE02A, and a stable algorithm for revising a triangular factorization of a symmetric matrix, MC11A. John Reid had arrived also, and worked with Alan Curtis on the solution of large, sparse systems of linear equations, which gave subroutine MA18A in 1971(5). Subsequently, John Reid extended this work to linear programming calculations. One of my interests in 1972 was the approximation of smooth curves by straight line segments for the purpose of graph plotting which, with the help of Sid Marlow, yielded OB11A and OB12A. At this time Roger Fletcher was testing some of his highly promising ideas for taking account of non-linear constraints in optimization calculations, which gave subroutine VF01A in 1973. Also in this year two important additions were made by visitors: Bert Buckley provided

VE05A[7], which is a substantial improvement over VE01A, partly because it can take advantage of any sparsity in the coefficients of the constraints, and Kaj Madsen contributed an algorithm that calculates all the roots of a polynomial.

Many other algorithms were added to the library in this period. Most of these came from writing Fortran programs ourselves to implement published descriptions of the methods of other researchers. A few subroutines, however, came from published computer listings. For example, the adaptive quadrature routine QA05A is based on CADRE due to Carl de Boor[8]. Because assembling these routines, and making them easily available to computer users, required a great deal of hard work, I would like to record here that the success of the library has been helped greatly by Christian Puritz, Bill Hart, Mike Hopper and Sid Marlow.

Because it is not possible for a group of numerical analysts to produce a string of good algorithms, without giving attention to the underlying theory, the group has also pursued an active research programme. The theoretical work on optimization is discussed in Roger Fletcher's contribution to this report, except that he is too modest about the importance of his studies on non-linear constraints. The list of reports in the TP series gives a fair indication of the underlying work of the group; but a further comment must be made on the research on sparse matrix calculations. It is that this research is pioneering work of high quality in a field that is essential to the solution of many very large computer calculations. Research has continued vigorously in this subject under the leadership of John Reid, since the Numerical Analysis Group transferred to the Computer Science and Systems Division, much of the recent work being done by Iain Duff, who joined in 1975. Therefore, Harwell is now recognised generally as a centre of excellence for sparse matrix calculations.

It is instructive to relate the work that has been described to research in numerical analysis and mathematical software outside Harwell. The attention that we gave in the 1960's to making available general Fortran programs for the algorithms that we developed was most unusual. As a consequence, we received many requests for these programs from outside and now hundreds of copies of the Harwell library are distributed throughout the world. It is therefore fair to say that the library has made an important contribution to scientific research generally. I have often received letters from scientists, thanking me for my algorithms, and saying that Harwell subroutines have made some mathematical calculations possible that could not have been done by other available methods. In the last ten years, however, the situation has changed greatly. The new subject "mathematical software", which falls between numerical analysis and computer science, has grown up at an enormous rate to provide libraries of high quality computer programs for numerical algorithms. It has consumed thousands of man-hours, because of the difficulties of making the programs available for a range of computing machines - difficulties which we have not had to face at Harwell. Work in this direction in Britain has been very worthwhile, its focus being the N.A.G. Library in Oxford. Many universities and research establishments, including Harwell, have contributed to this library, and its use is widespread, due to the attention that has been given to the requirements of different computers. It seems, however, that we may be going back to the situation where very few numerical analysts provide general computer programs, because this work will be done mostly by mathematical software experts. Unfortunately these experts often give insufficient attention to the mathematical theory of an algorithm, because so many other factors are important to them. Therefore I expect a return to the situation where Harwell will be unusual in providing subroutines that bring recent major advances in numerical analysis to computer users.

1. A.R. Curtis, The Approximation of a Function of One Variable by Cubic Splines in "Numerical Approximation to Functions and Data", ed. J.G. Hayes (Athlone Press, London, 1970) p.26-42.

2. M.J.D. Powell, Curve Fitting by Cubic Splines in "Numerical Approximation to Functions and Data, ed. J.G. Hayes (Athlone Press, London, 1970) p.65-83.

3. M.J.D. Powell, A Fortran Subroutine for Solving Systems of Non-Linear Algebraic Equations in "Numerical Methods for Non-Linear Algebraic Equations", ed. P. Rabinowitz (Gordon and Breach, 1970) p.115-161.

4. R. Fletcher, A General Quadratic Programming Algorithm, J. Inst. Maths. Applics, 7, 76-91 (1971).

5. A.R. Curtis and J.K. Reid, The Solution of Large Sparse Unsymmetric Systems of Linear Equations, J. Inst. Maths. Applics, 8, 344-353 (1971).

6. R. Fletcher, An Ideal Penalty Function for Constrained Optimization, J. Inst. Maths. Applics, 15, 319-342 (1975).

7. A. Buckley, An Alternate Implementation of Goldfarb's Minimization Algorithm, Math. Programming, 8, 207-231 (1975).

8. C. de Boor, On Writing an Automatic Integration Algorithm in "Mathematical Software, ed. J.R. Rice (Academic Press, New York, 1971) p.201-209.

Chapter XIX

OPTIMIZATION METHODS

R. Fletcher

I became interested in optimization methods in about 1961 whilst researching into numerical methods for computing atomic and molecular wave-functions. By this time the growing power of computers was making it apparent that many important optimization problems in industry, science and engineering might be solved numerically. This situation led to an increasingly rapid and fruitful study into numerical methods for optimization problems and into the theory associated with these methods, which has continued up to the present. What follows is a personal account of the many developments which have taken place.

The main centres for this development in the U.K. have been at Harwell, the N.P.L. and the Numerical Optimization Centre (N.O.C.) at Hatfield, together with a number of people scattered around the country in universities. Other contributions have come from operations research groups such as Scicon. New developments from industry have been few, although its representatives have attended enthusiastically at conferences, teaching symposia and summer schools, so that new developments in methods have been successfully used in many practical applications. In fact, the traditional approach in the U.K. is such that practical experiments with methods have not been neglected in favour of theoretical studies, as has often happened in the U.S. for instance. As a consequence, I believe that developments in the U.K. have had a practical significance far in excess of the relative number of people working in the subject. At Harwell the establishment and widespread dissemination of a subroutine library, which includes a wide selection of up-to-date optimization routines, has greatly increased the impact which these developments have had. In all this research it is appropriate to mention particularly my former colleague at Harwell, Mike Powell, whose contributions to the subject have been unsurpassed.

In the late 1950's it was virtually impossible to find an efficient and reliable method to solve other than the most basic optimization problem. This meant that the first contributions came from engineers and scientists in industry, often in a very ad-hoc way, although these often exhibited considerable ingenuity. Early research often studied what are the most suitable theoretical properties for methods to possess. As an example, a general unconstrained minimization method might terminate at the solution if applied to a quadratic function. An apparent early success was to show how termination could be achieved merely by minimizing the objective function successively along certain lines. Unfortunately, the resulting method did not work as well as the ad-hoc method it was intended to replace! However, Mike Powell was able to show a better way of using this theoretical property and the resulting method was used successfully for a number of years(1). This episode illustrated to me very clearly the importance of integrating both theoretical and experimental studies.

The classical Newton method for minimization is the most direct way of using the properties of a quadratic function and can sometimes be used advantageously. However, the basic method is unreliable and requires the user to evaluate second derivatives, which can be a major disincentive to its use. On the other hand, the apparently optimal steepest descent method, which requires only first derivatives, is virtually worthless in practice. In 1962 Mike Powell and I independently came across a rather idiosyncratic report by Davidon$^{(2)}$ which nevertheless contained a most important idea, namely that of approximating the inverse second derivative matrix by a certain positive definite matrix which is updated on each iteration. When used with a line search for stability, many of the difficulties of Newton's method were resolved at a stroke. The resulting DFP method$^{(3)}$ has been used widely and successfully for many years. The method was an important breakthrough: we attended a meeting about that time at which the speakers were lamenting the difficulties of minimizing functions of ten variables (the phrase "curse of dimensionality" was often used) whereas the DFP method would readily solve test problems of one hundred variables. One of the most striking early applications of the method (amongst many) was in helping to compute optimal trajectories in the U.S. moon landing programme. Since that time innumerable researches into this type of method have been carried out. Two significant developments which are particularly worthy of mention arose partly as a result of research at Harwell. One was the introduction of a new updating formula$^{(4)}$, which proved even more effective than that used in the DFP method, and which forms the current standard method. The other development was motivated by a bet of one shilling that a proof of convergence for the DFP method would not be obtained in the foreseeable future, in view of the difficulties which were involved. This challenge strongly motivated Mike Powell to solve the problem$^{(5)}$ and he has since gone on to develop many fine results of this nature. Other important developments in these types of methods in the U.K. have taken place at the N.P.L. and the N.O.C.

Another early development with which I was fortunate to be associated was the conjugate gradient method(6). Although not comparable with the DFP method, it has the important advantage that no matrices need be stored. As a consequence, it could be applied to very large problems (1000's of variables) for which other methods could not be used. The method found important applications in many fields, for example in studies of defects in crystals by Mike Norgett at Harwell(7). As time passed many of the properties of this type of method were established, and it was apparent that a number of similar methods could be derived. Experimental testing on the usual range of test problems failed to detect any noticeable difference between them. However, the large problems for which the method was most suitable have certain symmetry properties which are not well represented in the standard test problems. Again Mike Powell has been able to show(8) that one particular conjugate gradient method is most advantageous when the symmetry is present, and substantial gains in efficiency are possible.

Perhaps 90% of non-linear optimization problems are non-linear least-squares problems, arising either from data fitting, or as an attempt to solve a non-linear system of equations. Many developments at Harwell have contributed to the current effectiveness of methods for such problems, although a data fitting package at Harwell was given the acronym FATAL, namely Fit Anything To Anything you Like, a wry comment on the dangers which an unsophisticated user might encounter. Most of the developments again arise by estimating the second derivative matrix, this time by linearizing the non-linear equations. Occasionally this approach fails to work well, even when modified in such a way as will, in theory, guarantee convergence. The reasons why this is so are known, but the best way of developing better methods is not as clear, and the whole area of modified Newton and Newton-like methods awaits some new initiative.

Many minimization problems are complicated by the presence of *constraints*, that is equations or inequalities which the variables are required to satisfy. The minimum value of the objective function subject to these restrictions is then required. A basic problem of this type is *linear programming*, in which both the objective and constraint functions are linear. A standard technique for handling such problems was set out in the 1940's and is still in common use today, albeit with some modifications to improve efficiency. Contributions to this work have taken place at Harwell by John Reid for solving large sparse problems which frequently arise in practice. Also Ian Cheshire has successfully adapted linear programming to be applicable to ship scheduling problems. It soon became clear that there are no major difficulties in handling linear constraint problems in general, although suitable software is not always readily available. I developed a quadratic programming (quadratic objective function/linear constraints) method(9) and subroutine at Harwell in 1970, and this has been well-used. Earlier quadratic programming methods relied on modifications to linear programming, whereas this method was typical of a trend towards what might be called an *active set method* and most current research now follows this approach.

More recent developments which have had an influence on methods for linear constraints have occurred in the field of numerical linear algebra, with particular regard to factorizing and updating matrices in a stable and efficient way. It is now realized that substantial loss of precision can occur unless the correct approach is used. Workers at N.P.L. in particular have been responsible for making these effects known, and Harwell too has been well to the fore in these developments. Software, incorporating the more satisfactory matrix handling techniques which have been developed, is gradually becoming available.

The most challenging type of optimization problem is non-linear programming in which both the objective and constraint functions may be nonlinear. Early ad-hoc developments used the idea of adding to the objective function a term which penalized violations or near violations of the constraint conditions. The problem could then be treated as in unconstrained optimization. Theoretical studies showed that if a sequence of unconstrained problems were solved, then convergence to the required solution could be obtained. In practice, however, the methods were only capable of providing solutions of low accuracy - a fact which it also became possible to account for theoretically as an illustration of ill-conditioning. In the late 1960's a better sequential penalty function formulation was found which avoided the ill-conditioning and it became possible to provide reliable software for solving non-linear programming problems. I was rash enough[10] to refer to this as being an 'ideal' penalty function, an assessment which is already overtaken by more recent developments. Another type of penalty function which has attracted some attention is the exact penalty function, i.e. one for which a single unconstrained minimization could be used, rather than a sequence. A quite early idea was to use an L_1 (sum of moduli) penalty term. This, however, causes difficulties because of the lack of differentiability (i.e. smoothness) of the resulting function. I was able to propose[11] a smooth exact penalty function, which, however, never became competitive in practice.

An alternative line of development for non-linear programming, dating from about 1960, has been to linearize the constraint functions and to solve a sequence of quadratic programming problems, including a correction for constraint curvature in the quadratic function. Mike Powell has recently revived this idea, with a line search to help overcome problems of reliability associated with the earlier methods, and he has shown that the method can work well in practice. I have been able to show how to incorporate these ideas with an exact L_1-penalty function so as to guarantee convergence, while retaining the other advantageous features of the methods. These recent developments are very interesting and promising and I think it likely that in the future they may supersede the use of sequential penalty functions.

Finally, another recent field of interest has been in non-differentiable optimization. Use of L_1-penalty terms, both in non-linear programming and in solving non-linear equations, is one example, but there are a number of other important applications. In the past such problems have largely been avoided, due to both theoretical and practical difficulties. However, a number of successful research contributions on both fronts have been made, and it is likely that good practical software for such problems will become available in the not too distant future.

1. M.J.D. Powell, "An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives", Computer J., 7, 155-162 (1964).

2. W.C. Davidon, "Variable Metric Method for Minimization", AEC Res. and Dev. report ANL-5990 (Rev.) (1959).

3. R. Fletcher and M.J.D. Powell, A Rapidly Convergent Descent Method for Minimization, Computer J., 6, 163-168 (1963).

4. R. Fletcher, A New Approach to Variable Metric Algorithms, Computer J., 13, 317-322 (1970).

5. M.J.D. Powell, On the Convergence of the Variable Metric Algorithm, J. Inst. Maths. Applns., 7, 21-36 (1971).

6. R. Fletcher and C.M. Reeves, Function Minimization by Conjugate Gradients, Computer J., 7, 149-154 (1964).

7. M.J. Norgett and R. Fletcher, Fast Matrix Methods for Calculating the Relaxation About Defects in Crystals, J. Phys. C, 3, L190 (1970).

8. M.J.D. Powell, Restart Procedures for the Conjugate Gradient Method, Math. Prog., 12, 241-254 (1977).

9. R. Fletcher, A General Quadratic Programming Algorithm, J. Inst. Maths. Applns., 7, 76-91 (1971).

10. R. Fletcher, An Ideal Penalty Function for Constrained Optimization, J. Inst. Maths. Applns., 15, 319-342 (1975).

11. R. Fletcher, An Exact Penalty Function for Nonlinear Programming with Inequalities, Math. Prog., 5, 127-150 (1973).

Chapter XX

OPERATIONS RESEARCH AT HARWELL

Ian M. Cheshire

With the passing of the Science and Technology Act in 1965, the Authority was given the opportunity of making its research capabilities available to other organisations on a commercial basis. This is a particularly challenging task for theoretical physicists who are primarily concerned with explaining physical phenomena rather than exploiting their commercial potential. The tools of the trade are paper and pencil and the ability to make mathematical models. When the models become sufficiently complex computers are used to carry out the calculations. The application of these skills to commercial decision making problems is the prime concern of operations research.

In a typical operations research study we attempt to analyse a business problem and to construct as simple a model as possible which contains all the essential features of the operation. Depending on the nature of the operation the model may or may not be a simple one. If it is very complex or if it is stochastic in nature then we may have to resort to simulation on a computer with the aim of deducing rules of thumb to enhance the efficiency of the operation. Defining the design parameters of a new combat aircraft by simulating its performance under battle conditions would be an example of this type. However, sometimes the model can be simplified sufficiently to define a mathematical programming problem and we then have some hope of finding the optimum solution. The day-to-day scheduling of a fleet of road vehicles would be an example of this type.

At Harwell we have specialised in the development of computer programs to solve these extreme types of problems and we have largely avoided the middle ground where the problems are often too complex to yield an effective solution. However, this has not always been the case. In one of our early projects we attempted to optimise the operation of an entire paper mill instead of concentrating on one or two key aspects of the operation and allowing thecompany expert to define the external parameters. Our idea was simple enough. We thought that by simulating a sequence of orders through a series of machines we could evaluate many possible sequences and arrive at a good sub-optimal solution. The points we failed to recognise were (i) that the operation of each machine is always much more complicated than even the experts can tell initially and (ii) that people are often much better planners than we might suspect. In fact, our paper mill project was a failure, but from it we learned the virtues of simplicity and optimality.

Scheduling fleets of bulk carriers and OBO (oil/bulk/ore) vessels is an example of an operations research application where the central problem can be simply stated and where a globally optimal solution can be found. The uncertainties of the shipping industry are notorious and during the 1960's fleet operators attempted to shield themselves from fluctuating market conditions by signing long-term contracts at modest freight rates. For example, a company might contract to ship say 3 million tons of coal per annum from the U.S.A. to Japan. By using flexible OBO vessels they could then supplement the base trade by shipping oil from the Persian Gulf to Europe. This strategy combined two long laden voyages with two relatively short ballast voyages instead of the more conventional practice of devoting specialised vessels to a single trade and wasting half the sailing time in ballast. Seabridge Shipping Ltd., a large U.K. consortium, were prominent in this business but by 1968 the company had signed so many contracts that they no longer had a simple pattern of trade and scheduling the fleet became a major problem. Within about eighteen months or so we were able to write a computer program to schedule the Seabridge fleet in time to catch a peak in the spot market early in 1970. The market rate for shipping oil from the Persian Gulf to Europe had moved from $2.70 per ton to $27.00 per ton. In fact, by this time all the vessels were committed to much less profitable long-term business and the company scheduler could not see how to reschedule the fleet to free a vessel for the highly profitable oil trade. However, our program produced a "surprising" solution which freed vessels for the Persian Gulf to Japan route and the first computer run revealed a profit in excess of £1M. During the 1970's several other shipping companies made use of the program but the environment for which it was designed (the combination of long-term contracts and a profitable spot market) has since disappeared. At present, only one Norwegian company, specialising in the still buoyant business of shipping cars in bulk, continues to use the program regularly.

The central algorithm of the fleet scheduling program involves dynamic programming, to sequence the voyages of each vessel, as an interactive column generator for the linear integer program which allocates voyages to individual vessels. Integrality is achieved by a simple branch-and-bound procedure which either forces or denies particular voyages to particular vessels. The simplex multipliers or shadow prices change at each simplex iteration to define a new column generation sub-problem. The key to the solution of the sub-problem was the observation that discharge ports can always be grouped into a relatively small number of discharge areas and the dynamic programming stage-state can be defined as a discharge area-day.

Data validation is a fundamental aspect of any computer method but it is especially acute in real-time decision-making problems, such as scheduling a large fleet of ships, where one mistake can be very expensive. The normal procedure to cope with this data problem is to expose the model in as much detail as possible to enable the user to apply as many checks as he feels necessary to gain confidence in the model. For example, we would normally print tables of the cargo tonnage, etc., for each vessel on each potential voyage. Large sections of data which are relatively static such as contract details, distance tables and vessel characteristics can then be segregated into standard files. The remaining problem, which is not so easily solved, is how to verify the data, such as voyage laydays or market freight rates, which is constantly fluctuating. For example, by typing in the wrong month for a voyage layday, the operator may exclude the obvious schedule from the feasible set. Our solution to this problem was to exploit the schedulers' skills by requiring him to provide an input schedule which could be evaluated and checked for feasibility before allowing the program to search for the optimum schedule. Not only does this simple device eliminate errors in temporary data, but it also serves as a constant reminder to the user of the value of computer optimisation!

Scheduling fleets of road vehicles is a related problem with a wide market which has attracted many publications over the last ten years or so. Unfortunately, the problem is difficult and we are forced to use heuristic procedures which do not guarantee optimality. A recent study by the

Operations Research Group has shown that existing methods are capable of considerable improvement and we have therefore developed a new package to solve this problem in conjunction with Ross Foods Ltd, who have applied the technique with considerable success to a large number of depots. The marketing of the vehicle-routing package to a group of companies is now in progress.

Rather than day-to-day scheduling the main application of vehicle routing is strategic. It is usually possible to define a weekly cycle of customer demand and the problem is to service the customers from each depot with the minimum number of vans. Typical van costs are about £15,000 per annum and if a few vans can be saved at each of a large number of depots then the application of vehicle routing can become economic.

During the last few years the main interest of the Operations Research Group has centred on the development of mathematical techniques for simulating North Sea oil fields. This project is sponsored jointly by the Department of Energy, the British Gas Corporation and the British National Oil Corporation.

Computer models which simulate the movement of oil, water and gas through the porous rock of the reservoir provide one of the most useful ways of evaluating alternative development strategies. Such models incorporate all our knowledge about the geological structure of the field, the porosity and permeability of the reservoir rock, the location of wells, the physical properties of the oil, water and gas, etc. Initially there is a great deal of uncertainty about the data but, as the field is developed, we gain new information which enables us to improve the accuracy of the model. The process is called "history matching". Parameters of the model, such as rock permeability, are adjusted until there is close agreement between the computer calculation and the observed production and pressure at each well.

From the modelling standpoint there are two important features of North Sea fields:

(1) The reservoirs are large. But large fields with many wells require large mathematical models if they are to be modelled with any degree of accuracy. Large mathematical models are expensive to run on the computer and we need many runs before a good history match is achieved. There is therefore a need for highly efficient numerical methods to be developed. Our technique, which has turned out to be extremely efficient, uses incomplete Choleski decomposition in conjunction with a variant of the conjugate-gradient algorithm to solve the large sets of sparse linear equations arising at each time-step of the simulation.

(2) North Sea oil is light and highly mobile but under-saturated with dissolved gas. Thus as oil is extracted, and the field pressure reduced, gas does not come out of solution to maintain the field pressure and keep the oil flowing. In the North Sea, the oil must be driven by injecting water round the periphery of the field. Because the oil is highly mobile there is a sharp interface between the oil and the water and, if the extraction is carefully managed, we can get an almost piston-like displacement of the oil and a high ultimate recovery. However, there is no existing mathematical technique whereby this sharp piston-like displacement can be modelled accurately. Our approach has been to attack the problem in two ways. Firstly, by

exploring in depth the suitability of the finite-element method to reservoir simulation. This work is carried out wholly within the Theory of Fluids Group of Theoretical Physics Division. Secondly, by investigating new methods for incorporating sharp discontinuities in finite difference methods. Both techniques look promising and the latter is currently being incorporated into our comprehensive simulator called PORES (Program for Oil Reservoir Simulation).

The mechanism of oil displacement within the reservoir can be very complex and it would be useful to develop techniques to visualise the results of large simulation calculations. Some early work on this idea has been carried out at Shell using very expensive equipment. On the other hand, our idea is to use very cheap microcomputers (about £2K each) which can be used as private desk-top terminals by the reservoir engineers. This work is also at an early development stage but the results achieved to date have greatly impressed the practising engineers, who, in the past, have been forced to manage with much less effective methods.

The development of new mathematical techniques is, of course, only a means to an end and the results of our efforts would be of limited value if we could not use the new programs to carry out actual North Sea simulations on behalf of the Department. Having established a strong credibility for the Authority in this field, we could then propose a new programme to build individual models of North Sea fields. This part of the work is carried out at Winfrith and has turned out to be another sucess. Expertise at Winfrith grew rapidly during 1978 and that Establishment is now initiating new programmes on enhanced oil recovery methods. The Winfrith effort is now comparable in size with that at Harwell but it is growing much faster.

Chapter XXI

COMPUTING AT HARWELL 1948-1961

Jack Howlett

1. The Early Days

I came to Harwell in the summer of 1948 at the invitation of Dr. Klaus Fuchs. During the war I had been a member of a small group headed by Hartree, hidden away in a basement in the University of Manchester and working on a variety of mathematical problems connected with war-time activities. Our main tool for numerical work was the mechanical differential analyser, a fearsome and massive piece of mechanical engineering looking very much like a large-scale Meccano model - Hartree had in fact built a small prototype in Meccano before getting the full-scale machine. It was the most powerful calculating engine in Britain, probably in Europe, at the time - and an almost exact copy was built for Cambridge shortly after the Manchester machine came into use. It was, of course, an analogue, not a digital machine (although I don't think the name digital machine had been invented then). Using it was what one can fairly call man's work - you put on a boiler suit to change the set-up from one problem to another. One half of the Manchester machine is in the Computer Gallery in the Science Museum in South Kensington; I appear in the accompanying photograph.

We - meaning Hartree's group - had done what at the time was rated a large-scale calculation for Fuchs and Peierls, concerned with the atomic bomb project. Fuchs went to Harwell when it was set up in 1946 to form and head the Theoretical Physics Division, and about a year later asked if I would join his Division to take charge of the computing section he was building. Thus I arrived on the site in mid-1948 - 31 years ago.

Computing then was a completely different world from what it is now. There were no computers as such apart from tremendous adventures going on in a small number of institutions in Britain and America, amongst which the universities of Cambridge and Manchester were outstanding. The computing at Harwell, as elsewhere, was done with mechanical desk machines, either hand or electrically operated. When I joined the group consisted of about eight young people, mostly girls. Harwell was then just beginning its period of rapid build-up; the BEPO reactor had started up that summer and other large projects were under way. Fuchs had formed his computing group solely to serve T.P. Division but several things soon became very clear: (i) computational services were needed all over the Establishment, (ii) these needs were going to grow rapidly and (iii) what was wanted was a service available to everyone. It may seem strange today that an enterprise so large and so technologically and scientifically sophisticated should have been planned without a central computing service, but remember, this was 30 years ago and the power and all-pervading nature of computation could not possibly have been realised at the time. Anyhow, I told Fuchs of my views and he agreed; we then went to Sir John Cockcroft, the Director, with the proposal that the Theoretical Physics computing group should be developed into a station-wide service and he agreed; and that was the start of the Harwell computing organisation.

This was before the U.K.A.E.A. had been created, when Harwell was still part of the Ministry of Supply. The administration in London - who remembers the Adelphi? - classed desk machines with typewriters and other office equipment and controlled the supply rigidly. I made pilgrimages to the O. & M. people in the Adelphi to argue for almost every new machine we needed. I have a clear and pleasant memory of the occasion when, one morning, I convinced the head of that section - a very hearty Austrlian called MacPherson - that we really did need half-a-dozen new electric machines; and of celebrating the achievement by taking myself to lunch at Simpson's, more or less next door - and spending almost ten shillings on this blow-out. Official subsistence was probably about 3/6d. at the time.

There is scope for a tremendous amount of technique in hand computing and with good techniques and good organisation a surprising amount can be achieved. But the limitations are very great indeed and put a premium on carrying the mathematical development of a problem as far as possible before turning to numerical methods. This is especially true of what was the main field of work at Harwell at the time, nuclear reactor theory, where one is dealing with partial differential equations or, worse, integro-differential equations. Direct numerical attack was quite out of the question then and a whole battery of analytical weapons had been developed to make it possible to get anywhere at all with the limited computational resources available.* This, of course, applied everywhere, not just to Harwell. The late Boris Davison was an outstanding classical mathematician and a great expert here. He wrote, jointly with John Sykes (who is now with Oxford University Press) the definitive book on the subject, Neutron Transport Theory (Oxford

* An excellent and comprehensive account of the body of mathematical methods developed for the attack on these problems is contained in Nuclear Reactor Theory: Proceedings of the 11th Symposium in Applied Mathematics of the American Mathematical Society, April 1959, published by the A.M.S., 1961.

University Press, 1957), a tour de force of applied complex-variable theory. Boris knew his way around the complex plane better than anyone else I have ever met. The methods almost always led to some form of series expansion which was evaluated numerically term-by-term; doubtless many will remember the names "P_3" or "P_5" method, the suffix giving the order of the spherical harmonic to which the expansion was taken. Sheer algebraic complexity usually set the limit to this. Reactors being more or less cylindrical, Bessel functions turned up in most calculations and we made great use of published tables. The best tables by far were those published by the British Association; but we were still suffering severely from war-time restrictions and just could not buy the number of copies we needed. Entirely illegally, I got photographic copies made (no Xerox machines then). It's an interesting comment on the times, when one recalls that Harwell had pretty well unlimited funds.

Everyone who runs a computing service knows that when someone comes with a numerical problem the first thing to do is to find out if this is really the problem he wants solved, or whether it is, in fact, some transformation, which may or may not be useful, of the original problem. We had any number of experiences of scientists doing a lot of mathematics on a problem before bringing it to us, when a straight numerical attack from the beginning was far more effective. One case I remember is being asked to evaluate a singularly horrible mess of series, polynomials and quadratures which would have taken days of work; asking where it all came from I was shown a fairly simple non-linear ordinary differential equation and one of the girls got the result which was wanted by direct numerical integration in half a day.

2. The Punch-Card Era

Somewhere around 1952 we were asked by Dr. T.E. Cranshaw if we could help with what one would now call a data-collection problem. He was studying cosmic-ray showers and had an array of detectors on the Culham airfield; he needed to know, over a long period of time, which detectors had fired when, and to do a good deal of fairly simple arithmetic on the observations. This seemed a good application for standard punched-card accounting machinery and we went with the problem to the relevant manufacturers, British Tabulating Machine Co. (B.T.M.) and to I.B.M., the latter having just set up in a small way in Britain following the dissolution of the link between the two companies. Ironically, I.B.M. were not interested but B.T.M. were; they were enthusiastic and gave us a lot of help. We finished up with an 80-column card punch into which the signal cables from the detectors came and which punched a card showing the pattern of firings, and the time, whenever any one or a combination fired.

The reason for relating this is that the undertaking gave us elementary but valuable experience of punched-card machinery. Not long afterwards news of the Monte Carlo method of tackling neutron transport problems began to come over from America. Essentially a simulation method, in which one followed the life-histories of individual neutrons, this side-stepped the formal mathematical difficulties of the classical methods but required very large amounts of relatively simple computation if results of acceptable precision were to be obtained. Hand computing was quite inadequate but the method was well adapted to punched card machinery which was available as standard, commercially-produced equipment. Further, there was a good deal of experience of the use of this machinery for scientific computation, in the National Physical Laboratory especially. We set up a Punched Card Machine section around 1953 and lured James Hailstone (and his wife Elizabeth) from N.P.L. to run it.

We did a lot of Monte Carlo work with this machinery. K.W. Morton (now Professor of Applied Mathematics at Reading) had joined the group not long before and did a great deal to improve the statistical techniques and so to reduce the amount of computation needed in a problem. We did much general computation also, and here the Hailstones, who had a true flair for exploiting the capabilities of these machines, were invaluable. The manufacturers produced increasingly powerful and sophisticated machines capable of doing, automatically, increasingly complicated calculations, and we finally installed two of what I think can fairly be called the culmination of the punched card machines, the BTM.555. This isn't the place to give a long account of what I still think was a remarkable machine; so let me just say that whilst it was "programmed" by setting up a plug board, it had a magnetic drum store, allowed the repetition of program steps (DO loops?) and, in effect, the incorporation of sub-routines and could be made to do remarkable tricks. It was, in fact, not far off a computer. James Hailstone exploited this machine to the full; he wrote a short book, which B.T.M. published, describing the machine from the point of view of one concerned with scientific computation and giving half-a-dozen examples of applications: one was Monte Carlo, another a tricky combinatorial problem which arose in nuclear structure theory. I still have a copy.

I think it worth recording that we used the 555 for what must have been one of the earliest examples of automatic data processing. Reactor Physics

Division had a time-of-flight neutron spectrometer in the BEPO hangar for measurement of cross-sections. It was a simple but tedious calculation to go from the numbers they recorded in the experiments to the cross-sections they wanted, and they were doing this by hand and getting swamped. We automated the process, but not without difficulty. The basic calculation was simple enough; the real problem was to find out exactly what calibration and other corrections had to be applied to the raw observations, how these varied from day to day (as they did) and what other folk-lore came into the process. It was quite a salutary experience for both sides at the time.

3. Home-made Computers

The punched-card machine installation was in operation and doing good work from about 1953 to about 1957, but meanwhile the development of the true digital electronic computer was gathering momentum. Wilkes at Cambridge was building EDSAC-1, to be followed by EDSAC-2; Williams and Kilburn at Manchester were building the machine which the Ferranti Mark I was based on, to be followed by Mark I*; and at N.P.L. a group which included Turing and Wilkinson was building ACE, or more correctly Pilot ACE, on which the English Electric DEUCE was based. I went to what must have been the very first programming course in Britain, at Cambridge in 1950, and shortly afterwards some of the members of the computing group went to courses at Manchester. I and some of my colleagues went regularly to the seminars which Wilkes arranged on alternate Thursday afternoons in the Cambridge laboratory, as did members of the Electronics Division, notably E.H. Cooke-Yarborough, R.C.M. Barnes and D.J.H. Thomas. When petrol became more readily available, Cooke-Yarborough drove us there and back in his red Allard - a great experience. I've often told people that for several years the entire British computer population could and did meet together in the Cambridge lecture room.

It was becoming clear that the computer was going to happen and to be important, though I doubt if anyone foresaw just how important. There are, by the way, plenty of stories about estimates made at the time - for example that five or six machines at most would meet all foreseeable needs in Britain. These stories are true - I was at one quite high-level meeting at which such an estimate was agreed. No-one should laugh; who, only a few years ago, would have predicted the market for hand electronic calculators? However, as a step on the way Electronics Division offered to design and build for us an automatic calculator in which the switching was done by relays (as had been done in a classic series of machines built by Stibitz in Bell Labs.) but in which the decimal arithmetic and memory were electronic, using about 800 scale-of-ten Dekatron tubes. We got this in 1951 and housed it in the old control tower on the south-east corner of the airfield. It was, of course, slow, not much faster than hand calculation on single operations, but fully automatic, extremely reliable and utterly relentless.* It took little power and could be left unattended for long periods; I think the record was over one Christmas-New Year holiday when it was all by itself, with miles of input data on punched tape to keep it happy, for at least ten days and was still ticking away when we came back. It was perhaps only just a computer, but granted that, it was certainly one of the earliest

* One day E.B. Fossey, an excellent hand-computer (still with what used to be called the Atlas Laboratory), settled down beside the machine with his desk machine and attempted a race. He kept level for about half an hour, working flat out, but had to retire, exhausted; the machine just ploughed on.

in serious and regular use in the country. There is an account in Bowden's "Faster Than Thought". The subsequent history of this machine is interesting. We used it up to about 1958 and then, rather than scrap it, offered it as a prize to the educational institute which could give us the best reason for having it. This was the idea of J.M. Hammersley of Oxford and we conducted the operation in collaboration with the Oxford Extra-Mural Department. It went to Wolverhampton Polytechnic who used it for teaching and for real work for at least the next 15 years - an astonishing record; it is now in a museum in Birmingham, and I'm told it can still be made to work.

The first computers, and those in service right up to the early 1960's, were valve machines. Transistors began to appear in the early 1950's and Electronics Division immediately began to take an interest in their possibilities. In 1953 Sir John Cockcroft encouraged them to design and build a computer for us, using transistors throughout. The resulting machine was called CADET - Transistor Electronic Digital Computer (backwards). It went into regular service in an experimental form in August 1956. However, CADET used point-contact transistors which were the only ones available when the project started. But by 1958 these had been made obsolete by the development of junction transistors, so the machine was never re-built in a fully engineered form. It continued in regular use for about four years.

4. The Computer Era Starts

The first half of the 1950's saw the foundations being laid for the computer industry and also for methods for direct numerical solution of field problems, particularly the equations of neutron transport and diffusion. The second is, of course, not independent of the first; these methods lead to very heavy computation and there is no great incentive to carry their development very far if there is no possibility of actually using them. The greatly improved understanding of what was going on in these numerical processes led to the development of effective and efficient computer programs for studying, for example, criticality conditions in reactor assemblies. A.W.R.E. Aldermaston had installed a Ferranti Mark I* in 1954. We were able to get time on it, mostly to run a reactor-criticality program written by A. Hassitt of T.P. Division. A single run could take from 10 to 30 minutes and the reactor design people always wanted a lot of runs so as to explore some parameter space. We got most of our time at night, between about 10 p.m. and 6 a.m. the following morning. There was no such thing as multi-programming - you had the machine to yourself and you drove it from the console - so you were really working all the time. I, and most other members of the computing group, spent a lot of our nights in the Aldermaston machine room. Driving back at dawn over the Berkshire countryside was often quite delightful.

By 1955 it was obvious that we needed a powerful machine of our own at Harwell - "powerful" must be understood as relating to what was considered powerful at the time. Neither of the machines which were available then, the Mark I* and the English Electric DEUCE, seemed suitable. In 1956 Ferranti announced their Mercury, one of the first machines to have built-in floating-point arithmetic, which seemed very suitable. Mercury, like Mark I, was essentially a Manchester University design. As well as the novelty of floating-point it had a core store; it's wryly amusing to recall that Ferranti's first announcements of the machine spoke of the "giant" immediate-access store - 1024 words (of 40 bits each). We ordered one in 1956, not without some opposition from, curiously enough, parts of the scientific population; not everyone was convinced that a computer was really a necessity. The actual process of getting the machine ordered is a delight to recall. I had convinced my Division Head - Brian Flowers - (now Lord Flowers) that this was what we needed and he told me to go ahead. I wrote a one-page letter to Tom LeCren, who was then Secretary of Harwell, setting out the case and saying it would cost £80,000. He asked me to explain a few points and substantiate a few statements, accepted what I said and sent off the order. We got the machine in 1958 and installed it in decidedly slummy conditions in Building 328. Ferranti made about twenty of these machines, quite a number of them going to nuclear-energy centres; ours was, I think, number 4; there was already one at Saclay when we got ours, and later ones went to Risley and Winfrith and also to CERN.

All early computer users wrote their programs in machine code so programming was something of a black art. R.A. Brooker in Manchester (now Professor and pro-Vice Chancellor at Essex) devised an "Autocode" for the Mark-I and Mark-I*, which was a simple, easy-to-learn and easy-to-use high level language - not very high, to be sure, and very slow in execution, but a great improvement on machine code if you had only a fairly small program to write. We introduced this to Harwell and it caught on. When Ferranti embarked on building Mercury - based, as I said, on the next Manchester design - Brooker decided to write a new Autocode which, because of the higher basic speed and better facilities offered by Mercury, would be much more powerful and flexible and much faster. He spent a lot of time with us at Harwell discussing what should go into the new system, so that Harwell certainly had a significant influence on what he produced. The system, Mercury Autocode, proved a great success. Although simple when compared with modern high-level languages it provided an admirable range of facilities and was very easy to learn. With various enhancements it had a remarkably long life; a compiler for the final version, called CHLF because it was produced by a collaboration between CERN, Harwell, London (University) and Farnborough (RAE), was written for Atlas and was still in use in the early 1970's.

5. The Atlas Project

Things began to move very fast in the computer world in the second half of the 1950's. Technology, particularly that of producing core stores, improved greatly and several American companies began to produce bigger and more powerful machines than Mercury; above all, I.B.M. entered the field in earnest with the 704. This succeeded their first large-scale machine, the 701, of which they had produced eighteen between 1952 and 1954. It had built-in floating point, like Mercury, but was altogether on a larger scale and, in particular, could have what was then a scarcely believable size of core store, 32 K words. It is only fair to add that it cost a very great deal more than Mercury. A.W.R.E. installed one of these machines in 1957 and we used it for the increasingly large amount of work which was too big for Mercury. Tony Hassitt had rewritten his reactor program for Mercury and the new version, much faster and more powerful, was much used. But the spread of these bigger machines in America had led to the development of bigger and more powerful programs, and of linked suites of programs, which the Harwell reactor engineers and physicists wanted to use. These programs were produced in the big American nuclear laboratories such as Argonne, Knolls (General Electric) and Bettis (Westinghouse); also much important work was done at Los Alamos, including, for example, the production by Bengt Carlson of the simple but ingenious S_n method for direct numerical integration of the neutron transport equation. I think it was about this time that the writers began to give names, mostly acronyms, to their products.

The 704, like Mercury, was a valve machine. Towards the end of the decade I.B.M. produced a transistor (and enhanced) version, the 7090, which was probably the first large-scale machine using transistor circuitry. A.W.R.E. replaced their 704 by the faster 7090 in 1960.

As early as 1959 I and several of my colleagues were feeling concerned at the way computer production was going. We could see that at least in the scientific and technological field (and especially the reactor field) there was a need for ever more powerful machines. American industry was already clearly in the lead and I.B.M. was embarking on the Stretch machine which at the time seemed to be aiming almost at the ultimate in computers. Talks between myself, Bill Morton, Ted Cooke-Yarborough and John Corner (in charge of computing and theoretical work at A.W.R.E.) led to the view that something should be done to get an advanced, big machine project going in Britain. Corner and I wrote to Sir John Cockcroft; he was most sympathetic to the idea and wrote to all the leading people in the country concerned with computers inviting them to come to Harwell to discuss the needs, problems and possibilities. There was a ready response and we had two very constructive meetings in which, in addition to A.E.A. people, Wilkes, Williams, Kilburn, Strachey, Halsbury (then Managing Director of N.R.D.C.)

and others took part. Everyone agreed that there was a need for an advanced machine project to be supported in Britain. The final decision was that this should be based on the new machine then being designed at Manchester by Kilburn and his colleagues, a machine which satisfied the criteria of computing power, storage capacity and general flexibility and sophistication at which we had arrived. Whilst there were great and important differences between the two concepts, the aim was a machine in the same class as I.B.M.'s Stretch. A.W.R.E., incidentally, installed a Stretch in 1962.

Merely to shower blessings on an R. & D. project was a long way from being enough. What was important was that the Manchester design should become a properly engineered machine, manufactured by the computer industry and sold as an industrial product. In 1960 Ferranti, who had a long association with Manchester University, indicated that they would engineer and market the new design if they were guaranteed one order. They proposed to call it Atlas - after Mark I they had given all their machines mythological names: Mercury, Pegasus, Orion ... Sir John Cockcroft, who all along had taken the view that this was a project which the A.E.A. might well support, gave me the job of building up the case. I spent most of 1960 going round the A.E.A. Establishments (not A.W.R.E., understandably, although I had many talks with Corner) trying to assess the future demand for computing; Cooke-Yarborough helped a lot. The Authority accepted the case for buying an Atlas, but because of the cost, nearly £3M, they had to have the approval of the Treasury and of the Minister for Science, Lord Hailsham. After much discussion, in which Sir William (now Lord) Penney took an important part, Sir John Cockcroft having now left Harwell to become the first Master of Churchill College, Cambridge, the following proposal was put to the Treasury and to the Minister:

1. The A.E.A. should order an Atlas from Ferranti.
2. The machine should be set up at Harwell.
3. The power of this machine being so great that only half its time would be enough to meet all the needs of Harwell and the overflow needs of Culham, Risley and Winfrith, a substantial amount of time should be made available, as of right, to the newly-formed Rutherford Laboratory and to Universities generally.

The authorisation which came out of the Minister's office, however, was:

1. An Atlas should be ordered, as proposed.
2. The machine should be under the control, not of the A.E.A., but of the National Institute for Research in Nuclear Science (N.I.R.N.S.), the body set up to build and administer the Rutherford Laboratory.
3. Equal shares of the machine's time should be allocated to
   (a) Universities and Government organisations
   (b) N.I.R.N.S.
   (c) Harwell
4. N.I.R.N.S. and university users should not be charged for their use of the machine, other users should pay.

The last statement reflected the fact that N.I.R.N.S. had been set up with the right to provide services without charge to universities, whilst neither the A.E.A. nor Government bodies had this right.

N.I.R.N.S., with Lord Bridges as Chairman, then took over the project and agreed to build a new laboratory, the Atlas Computer Laboratory, on the Chilton site. The computer was ordered in the summer of 1961 and Ferranti gave a splendid lunch party at the Savoy in September to celebrate this. I recall saying to Tom Kilburn as we left that I should never have expected computing to run to such high life. I was given the job of Director of the new laboratory and transferred from the A.E.A. to N.I.R.N.S. in December to get this new enterprise going. My thirteen years with Harwell had been immensely exciting and enjoyable and so were the next fourteen with Atlas. But that is another story.

Chapter XXII

COMPUTING AND DATA PROCESSING

A. R. Curtis

1. Introduction

The time-span covered by the history of the Atomic Energy Authority virtually coincides with that of the development of modern computing. During the past 25 years, computer speeds have increased by a factor of nearly 10,000, while the cost per operation has decreased (in real terms) by a factor of more than 1,000; computer memory sizes have gone up by similar factors. Two major revolutions in computing hardware (valves to transistors to integrated circuits) have occurred, and more frequent developments of technology have made possible the enormous quantitative improvements mentioned above. On the software side, we have seen the development of ever more sophisticated operating systems, enabling the routine execution of 1000 or more computing jobs during an 8-hour daytime shift, compared with perhaps 5-20 dedicated individual user sessions 20-25 years ago, and the almost universal use of 'high-level programming languages' (e.g. Fortran). The problems solved have also become larger and more sophisticated, and certainly involve more realistic modelling of the external world; extensive developments in numerical analysis and other mathematical techniques, and in experimental validation of computed results, have played an important part here. Those not personally involved in computing have sometimes failed to apprehend the pace of development, while those who were involved have often failed to see the wood for the trees.

In retrospect, therefore, one can understand the considerable and increasing part which computing has played in the work of the A.E.A., and the complexity of attitudes towards this relatively new branch of technology. At one extreme, computing has been treated as almost an excrescence on the Authority's programme, to be contained in the face of pressures to increase expenditure on it without limit; at the other, the Authority's role has been seen as that of the only major U.K. user of large scientific computers, who must drag the indigenous computer manufacturing industry into the (late) twentieth century. Confusingly, both these extreme views, and many shades between, seem to have been held simultaneously, and perhaps this was an inevitable part of the process of coming to terms with such an upstart topic.

For example, because of the lack of a developed body of experience, A.E.A. assessments of computers generally over-estimated the overall computing power which would be developed in, say, three years' time by a revolutionary new computer system (whether proposed by a British or American manufacturer), in relation to the power of the 'production' computer (often American) then in use (or of some standard upgrading of it over the same time scale). In retrospect, this optimism was frequently due to lack of appreciation of the effort needed to provide efficient system software for the new system, and to get it working reliably; typically the new machine would approach its planned performance and reliability a year or two late. Recently, the advantages of stability have become better appreciated.

It is necessary, in the interests of brevity and reasonable accuracy, to find a formula for restricting the scope of this note to topics within the personal knowledge of its author. I shall deal, therefore, only with the use of digital electronic computers for scientific calculation and data processing, mainly from the point of view of the Research Group. The Authority-wide approach which has always been taken will, however, necessitate continual consideration of computing developments in the rest of the A.E.A.

I shall be concerned almost exclusively with two kinds of computer, (i) the large "main-frame", capable of supporting a mixed load of calculation and off-line data processing and (ii) the "mini-computer" (often called a "data processor"), dedicated to a particular task or related range of tasks, usually involving "real-time" processing of data from an experiment or instrument to which it is directly connected. The first such mini-computer was installed at Harwell in 1964; there are now about 90 of them, a few of which perform more sophisticated tasks, e.g. supervising the operations of a group of smaller ones. During the last two or three years, some of the earliest of these minis have been taken out of service after useful lives of 12-14 years. In contrast, it has been rather exceptional for main-frame computers to be kept in operation for more than about half this time. To a considerable extent, this is due to the enormous increase in the amount and variety of main-frame computing, together with a sharp decrease in the cost per unit of computing as improved technology has come into use.

We shall try to quantify the growth in computing later, but first it may be helpful to mention a few exceptions to the above simple classification into main-frames and mini-computers. In the late 1950's and early 1960's some main-frame computers did not have adequate manufacturer-supplied system programs which would have enabled them to interleave efficiently actual computation with the input of programs or data from a punched-card reader and the output of results to a line printer. Some, indeed, lacked even the hardware features to do this. Accordingly, a less expensive satellite computer was used to copy a batch of jobs from punched cards to magnetic tape, which was then used as input by the main-frame; the output was written on to another magnetic tape, which was then carried back to the satellite computer for copying to a line printer.

It was realised early on that the satellite could be remote from the main-frame and large computing loads were done at a distance, physically transporting magnetic tape by road, rail and air. When suitable equipment became available, data links were installed to copy reels of magnetic tape at a distance instead of physically transporting them, at least for the more urgent work. With advancing technology, it was later possible to use the data link as a direct connection between the main-frame and the satellite, which then became a remote job entry (R.J.E.) station. Early satellite computers in the A.E.A. were general-purpose machines intended for business data-processing, e.g. IBM 1401, and later IBM 360/30; British business computers then were not normally capable of handling the (IBM-standard) "industry-compatible" magnetic tape units. More recent R.J.E. stations tend to be special-purpose devices using mini-computers or micro-processors.

More recently, two different developments have appeared: the "large mini" and the "micro-processor". The latter is perhaps most usefully thought of as a programmable logic element for incorporation into a scientific instrument (or other piece of apparatus) to provide greater flexibility of control and use, without the user necessarily being aware how this flexibility is achieved. Within this restricted area, micro-processors (of which the A.E.A. has many) are not really part of the computing scene, except that people who design and implement such uses need similar skills to those in computing.

The large mini is perhaps more controversial, since it seems that it may erode the distinction between minis and main-frames, perhaps even replacing the latter by "distributed processors" - a network of mutually-coupled large (and small) minis, co-operating to carry out the necessary functions. Personally, I doubt whether this will happen to the extent that the main-frame user, or the installation manager, need concern himself over it. The advantages of going to a single manufacturer, experienced in developing and supplying hardware and software for a large installation, will continue; if such a manufacturer sees advantages in carrying the distributed processing approach further than is done at present, that is up to him. Already a modern main-frame includes many processors carrying out specialised functions, e.g. input/output.

On the other hand, an establishment normally has more than one computer installation. Those carrying out specialised functions, unsuitable for the mainframe, may well, in future, become large minis and so serve a wider range of users. In order to provide this wide service they may be connected to the same terminal network as the main-frame and have a direct data link to it. This seems to me a more likely development than replacing a powerful main-frame by a number of co-operating large minis.

2. Main-Frame Computers - To the End of the 60's

The needs of the weapons programme dominated the computing requirements of the A.E.A. in its early years. Altermaston had ordered a Ferranti Mark I* computer in September, 1953 and it was installed in April, 1955. In 1956 this was supplemented by an English Electric Deuce, and an IBM 704 was ordered. Both the Mark I* and the Deuce had small memories (of a few hundred "words" - each capable of holding a number) with magnetic drum 'backing stores' and were difficult to program (because of the need for absolute memory addressing). Nevertheless, much useful work was done on these early computers. The IBM 704 was delivered in February, 1957; it had 8K words (1K=1024) of memory and a drum, and was shortly afterward upgraded to 32K words (without the drum); it had 8 magnetic tape drives and a punched card reader, card punch and line printer. Moreover, it was programmed with the aid of a powerful 'symbolic assembler program', which removed from the programmer most of the need for absolute memory addressing. By 1959 a good compiler for Fortran, the new 'high level programming language', was available. This system was the start of a line of development which led directly to the modern main-frame and the way it is used; in retrospect it is regrettable that the early experience gained on it by A.E.A. staff could not have been spread more widely among the U.K. scientific computing community.

Harwell staff used Aldermaston computers (travelling by road to operate the computer themselves for agreed sessions of an hour or two) from 1956 and in that year a Ferranti Mercury was ordered for Harwell. This 'next generation' to the Mark 1 had a 1K-word memory and a 32K word magnetic drum, with punched paper tape as input/output medium; it had a modest assembler program and Autocode, an 'intermediate-level' programming language. (At the same time there was also a home-built transistorized computer called CADET at Harwell, but this was little used because it was difficult to program; see also Chapter XXI by Jack Howlett.)

Mercury came into full operation (after some delay) in the autumn of 1958, and was soon run on a three-shift basis. Two more were then ordered for civil R. & D. work, one for Risley and one for Winfrith. Meanwhile, this work continued to use Aldermaston computers, including the 704, whose computing power was at least three times that of Mercury, and which was being run on a four-shift basis (150-160 hours per week).

An IBM 709 replacement for the 704 was ordered early in 1958 and was delivered in mid-1959, when the 704 was moved to Risley. The 709 provided a moderate improvement in computing power because input/output transfers between memory and magnetic tape were no longer done by the main processor. But the main improvement was in software, the 'Fortran Monitor System' allowing 'job batch' operation via a satellite computer which did the card reading, card punching and line printing. Before delivery of the 709, IBM's five times faster,(but otherwise identical) 7090 was ordered as a replacement for it, for delivery in the autumn of 1960.

Also in 1959, a card reader and punch and a line printer were added (as a special development for the A.E.A.) to the Mercury at Harwell, and the Winfrith machine also had these additions when it was commissioned in 1960. Software support for this extension was done at Harwell, which maintained close contact with the Manchester University software development team for Mercury.

In 1960 an IBM Stretch computer (later known as IBM 7030) was ordered to replace the 7090 early in 1962 at Aldermaston. This was faster by job-dependent factors ranging from 1.5 to 5, but was also more powerful because of its much larger memory (96K words, of which 80K were available to the user and 16K were taken up with 'system' programs and buffers) and large, fast magnetic disk backing store. IBM supplied a 'Master Control Program' to supervise Stretch operations, but for some time the only acceptable Fortran compilers were the Aldermaston-written S1 and its faster replacement S2. Because civil work still depended heavily on the use of A.W.R.E. computers, Harwell provided assistance in the development of S2, and this came into use by June, 1963.

When Stretch was delivered in May, 1962, the 7090 was moved to Risley (thus replacing the 704) for civil work. At this time, there were IBM 1401 computers, acting as satellites to Stretch and the 7090, at Aldermaston, Culham, Harwell, Risley and Winfrith. As well as the three Mercuries, there was a Deuce at Capenhurst and an IBM 1620 at Windscale; of these the Mercuries, at least, were being run for 120 hours per week or more.

It is of interest to study the figures in Table 1, which contains forecasts made in January, 1963 of the growth of civil computing by the various establishments. These forecasts were drawn up paying some attention

	April 1963	April 1964	April 1965	April 1966
Harwell Culham Winfrith Risley	28 25 42 84	35 25 61 84	48 25 79 84	95 60 79 84-100
Total	179	205	236	318-334

Table 1
Forecasts of Civil Computing Needs (in 7090 hours/week)
Made in January, 1963

to constraints on the available capacity, which explains the large growth forecast by Harwell and Culham in 1965-66. The overall growth forecast by Harwell at that time, by a factor 3.4 over 3 years, was close to the exponential growth (with a doubling time of 1.8 years) which was later found to have characterized Harwell computing throughout the period from 1957 until at least 1969. Also in 1963 a comparison of computing costs was made by the Computer Policy Committee. The three Mercuries had together cost £470K to buy, and were costing £270K per year to run; while the 7090 (including the satellite 1401) which had cost £1380K to buy and was costing £490K per year to run, was at least fifteen times as powerful as a Mercury.

The middle 1960's were a somewhat troubled period in A.E.A. computing. The main themes were: (i) shortage of capacity for the immediate work-load (somewhat alleviated by a decline in weapons work), (ii) negotiations with Ferranti Ltd. over their new Atlas computer (comparable in power to Stretch), (iii) increasing pressure against centralized computing by those establishments having only satellites, (iv) concern over the growing cost of computing, (v) difficulties over Atlas hardware and software, (vi) increased political sensitivity of choice of computer and (vii) the announcement of new models by various manufacturers. Perhaps it is best to summarise the position in early 1967 without devoting too much space to the intermediate stages. By this time Atlas 1 was working at the Atlas Computer Laboratory adjacent to Harwell, using (among other programming aids) a symbolic assembler program and Fortran compiler developed by Harwell. Atlas 2 (a version without the famous 'paging' hardware) was working at Aldermaston alongside Stretch (which had now been purchased instead of rented), using (almost exclusively) the S3 Fortran compiler developed there. An IBM 360 model 65 was due to be delivered to Harwell in September, 1967 and was estimated to be 1.5 times as powerful as Stretch (after delivery tuning of the system software put this factor up to 2). Meanwhile, Harwell was not, as originally planned, using Atlas 1 largely because of cost and of pressure of other Atlas 1 users. Large configurations of the English Electric KDF9 computer (about equivalent to IBM 7090s) were working at Culham and Winfrith with software developed to the A.E.A.'s specification and with A.E.A. assistance. (The other KDF9 installations, in British universities, had only 8K word memories instead of 32K, used paper tape as input/output medium, and had no Fortran compiler). The IBM 7090 was in heavy use at Risley.

The A.E.A. first proposed in January, 1965 that the universities' KDF9 computers, referred to above, should be upgraded to match the two A.E.A. installations, as a relatively cheap way of disseminating computing experience and facilities, which had for years been available in the A.E.A., among the U.K. scientific computing community at large. Unfortunately the Universities' Computer Board was not able to put this imaginative proposal into effect until several years later.

Harwell's 360-65 was commissioned on time (and the 360-30 was relinquished) and by the end of 1967 Harwell had virtually completed the transfer of its workload from Aldermaston. (In mid-1967 a total of about 75 Stretch hours per week was sent back and forth by road on magnetic tape, using a satellite 360-30 at Harwell.) The load rapidly built up to bring Harwell's total back to its 1.8-year doubling time growth, and then continued at that rate.

From the start it had been intended that much of the work on the 360-65 would be entered 'on-line', via a network of teletype terminals and mini-computers, from all over the site, and - IBM's software for the purpose having proved over-ambitious - the Harwell User's Workshop (HUW) system was designed, to provide terminal users with file editing and job entry facilities with minimal overhead to the background load on the 360-65. This system came into use in 1968, but was plagued with reliability troubles for its first year of operation. Since then, it has made a valuable contribution to simplifying the use of the Harwell central computer. A majority of the 8,000 jobs per week done on Harwell's central computer are now put in via terminals.

3. Mini-Computers

Meanwhile, the mini-computer scene had become active at Harwell. It was realised that the kind of interactive real-time computer collection (and early analysis) of large volumes of experimental data which had always been envisaged was not practicable, for a number of technical reasons, under direct control of a general-purpose main-frame computer and this continues to be the case. (In fact, the Rutherford High Energy Laboratory had a dedicated main-frame type of computer, a Ferranti Orion, for data capture.) At the same time, the Digital Equipment Corporation (DEC) brought out their famous PDP-8 mini-computer, which was so competitive in price (in all its configurations) that it was relatively easy to demonstrate worthwhile savings by using it on such projects. Harwell installed mini-computers rapidly (attracting surprised comment from other establishments, some of which later followed suit). Among these were two 'message concentrators' to interface teletype terminals to the 360-65 with minimal load on the latter (one of these also provided some higher-speed data links to other mini-computers on site). By May, 1971 there were 23 mini-computers at Harwell, of which 15 were supplied by DEC; most of the others, mainly somewhat larger ones, were Honeywell DDP-516's.

To a considerable extent, the development of the use of mini-computers had been quite separate from that of main-frame computing; the latter had always been run by Theoretical Physics Division (which contained also mathematical support staff), although most divisions were users, while the former was done in the various experimental division, with Nuclear Physics and Materials Physics Divisions leading the way. Towards the end of 1973, after successful conclusion of a project of software assistance to a large U.K. company which had involved staff with both kinds of experience, the Computer Science and Systems Division was formed. Since then it has been responsible for mainframe operation and development, and has played a leading role in all aspects of its discipline, including the development of real-time systems and computer networks.

4. Later Developments

Since the late 1960's, the pattern of development could be simply described as 'more of the same'. Early on, Culham consciously decided that it wanted as much computing as it could get for a fixed fraction of its budget, and other establishments have effectively taken the same course. Harwell has upgraded its IBM 360-65 successively to a model 75, then model 370-165, then model 168 (which has 'paging' hardware, enabling 'virtual storage' operations), then improved the 168 to a 168-3, then upgraded it to a model 3033; a second 'attached processor' will soon nearly double the power of the latter.

All these upgrades, giving an overall computing power increase over the 360/65 by a factor in the range from 15 to 45, have been within a compatible range, and have scarcely affected the users; the uncertainty in the power ratio is mainly due to difficulty in estimating the effect of an increase in memory size from 0.5 MB (512K bytes, or 128K words), plus 1 MB of slower memory, on the 360-65 to 12 MB on the 3033. The introduction of virtual storage enables all but the very largest jobs to ignore the limitations of the actual memory available, while the large actual memory allows a high degree of multi-programming (so that the computer can be kept fully occupied), while permitting the more frequently used parts of the system software to be 'locked down' in actual memory, together with supervisors for the various terminal services supported. In addition to HUW, the latter now include scientific interactive systems TSO and APL, as well as others for purposes such as stock control by Stores. About 1,000 Harwell staff use the 3033 computer.

Meanwhile, Winfrith and Culham have replaced their KDF9's by English Electric (now ICL) 4-70's, each comparable to Stretch in power and similar to the IBM 360 series in instruction set; one of these was also installed at Risley to replace the 7090. The Risley 4-70 has itself recently been replaced by an ICL 2980, and other computers of ICL's new 2900 series are on order for Culham and Winfrith to replace the 4-70's there. The Authority is again short of computing capacity in the short term, and (especially with the needs of JET at Culham) the demand on Harwell's installation from other establishments continues to grow, in spite of earlier forecasts of a decrease.

The underlying reason for the increase in computing is that by its use one can save experimental costs, and can get results in shorter elapsed time. Computer modelling of experiments is now, in many cases, so good that, with care to normalise the results of computation against a few carefully-chosen experiments, it can be used for quantitative prediction of the kind needed to design new equipment, or to plan operating cycles, or to optimise performance. The uncertainties in comparing powers of different computers well separated in time, and of allowing properly for inflation, make it difficult to be too precise, but I estimate that the unit cost of main-frame computing has come down by about a factor of 1,000 (in real terms) at Harwell over the last 20 years. Over the same period, the real cost of experimental work has probably increased.

5. Some Achievements

It would not be proper to end without trying to give some idea of what has been achieved by all this computing. I shall not attempt a synoptic view because, at least at Harwell, computing activity is carried on by so many individuals in support of so many branches of science and technology. Instead, I select only a few examples with which I have been personally connected, over a time-span covering most of the 25-year period.

1. The early 1960's saw the completion of STAB, the first successful 3-dimensional simulation of a complete nuclear reactor. This program modelled the kinetics of thermal neutron flux with delayed neutron emitters, poisons, temperatures of fuel, coolant and moderator, with temperature-dependent multiplication factor and individual vertical movement of control rods in eight sectors driven by realistic three-term controllers based on

on coolant outlet temperature, for the Magnox reactor. It was used to study the non-axially-symmetric perturbations (only marginally damped because fixed absorber was distributed to improve mean power density) which could be excited by various kinds of accident. The program was written by a small team from Harwell and Winfrith, using symbolic assembly language, rather than Fortran, and occupied virtually the entire 32K word memory of the IBM 7090 computer in spite of 'packing' the numbers describing delayed neutron emitter populations three to a word.

2. The Atlas 1 computer would not have had a Fortran compiler, or even a symbolic assembly program of the kind we were used to, providing mnemonic operation and operand names, but for the Atlas Fortran project at Harwell. This was done by what is now known as "boot-strapping"; most of the code for the compiler was itself written in Fortran, running first as a cross-compiler on the IBM 7090 and thence being transferred to Atlas, with only utility sub-routines written in Atlas assembly code. The dialect of Fortran implemented also had several forward-looking features, which might with advantage have been more widely adopted: global variable and array names, for example, and the ability to change array dimensions by re-linking already compiled subroutines. (Although Harwell was not directly concerned, it is of interest to recall that, later, it was only the Authority's interest in Fortran which made it available on the English Electric KDF9 computers, and that Aldermaston produced the S3 Fortran compiler for Atlas 2. The high level languages developed by the manufacturers were, respectively, Atlas Autocode for Atlas and Algol for the KDF9, but these never in practice achieved sufficiently widespread acceptance to give comparable opportunities to Fortran for program portability and exchange.)

3. The HUW terminal system, carefully designed and implemented by Harwell to minimise demand on computer resources by terminal users, has provided a powerful yet easily-learned interface between users at Harwell and the I.B.M. central computer for a decade. During the first half of this time, no competitive system was available from I.B.M. and there are still some aspects of HUW which are superior to the corresponding features of T.S.O.

4. The FACSIMILE program, for simulating mass action kinetics (with diffusion and advection if desired) is now widely used both inside Harwell and elsewhere. It is based on Gear's introduction, about 1968, of methods of backward differentiation for stiff initial value problems, married with Harwell techniques for handling sparse systems of linear equations (and also for non-linear least-square fitting). The first successful test, eleven years ago, of the underlying numerical method ran 500 times as fast as previously-used methods, on a biochemical test problem (glycolysis) which seemed very large then (73 coupled differential equations) but would be classified as only of medium size today. Efficiency has since been further improved, but the popularity of FACSIMILE with its users is mainly due to the fact that it provides them with a specially-designed high-level programming language which they find suitable for describing their problems. It is used extensively by chemists, physicists, metallurgists, chemical engineers, reactor engineers, astrophysicists, biochemists and others, and has generated widespread confidence among its users in the results obtained. The largest problems solved with its aid have come from the Atmospheric Pollution Group at Harwell, involving over 1,000 simultaneous differential equations; one of their problems involved a time constant as short as 10^{-9} sec., with simulation required over a span of 10^{10} sec.

Finally, a word about computing staff. During the 25-year period, Harwell-trained staff provided cadres for the Culham and Winfrith laboratories of the A.E.A., and for the Atlas Computer Laboratory of the S.R.C. There are now at least three university professors, in mathematics and computing, who started their scientific careers at Harwell and stayed there (or at Culham) for many years, as well as less senior university academic staff.

Chapter XXIII

REAL-TIME COMPUTING

I. C. Pyle

1. Introduction

During the first two decades after the invention of modern computers (the 1940's and 1950's), their mode of use for both scientific and commercial purposes was as independent information handling machines. Computers were always isolated from their environment by layers of human activity. Developments in the 1960's led to a new mode of use, in which the computers were directly connected to other equipment, for instrumentation and control without human intervention. This mode has been called "on-line" or "real-time" computing, emphasising particular features of the connection; this kind of use is currently becoming described as 'embedded computing' to emphasise the relationship between the computer and its information-environment. This paper reviews the development of real-time computing, and explains how it differs from the conventional style of computing.

Real-time computing began in several different areas during the 1960's, in industry and in scientific laboratories, as the potential of computers to process data rapidly became apparent. At Harwell the major stimulus was the need in nuclear physics experiments to record results in large quantities, fast and accurately, with the experimenter able to monitor this data as it was acquired. In Theoretical Physics Division, we first met the problems in designing an interactive computing facility (HUW) which needed a 'front end' computer to communicate with both the terminals and the main I.B.M. system 360. The requirements for nuclear physics instrumentation led not only to developments in electronic standards (CAMAC [1]) but to experience in designing and programming real-time computer systems for control, data display and storage. Harwell was able to make a marked contribution to the development of Linesman, the British air traffic control system, as a result of this. The team worked with R.R.E. Malvern (now R.S.R.E.) and Plessey Radar to design and implement the software for the project in the late 1960's and early 1970's.

Most groups engaged in real-time computing during the 1960's were not aware of one another, and the unity of underlying ideas only began to emerge during the later part of the decade. As a result, the early systems were very difficult to design and maintain, with high costs and low reliability. The situation has been improving during the 1970's, with the emergence of some necessary design concepts and the commercial availability of appropriate system components. Most real-time systems are now designed to make use of equipment with standard interfaces and sometimes with standard software, although the software remains a serious problem. A major effort since 1975 to bring the benefit of software engineering to embedded computer systems has resulted in the recent publication of the programming language Ada [2], which can be expected to have a significant impact in the 1980's.

2. Economic Influences

Apart from the technical problems of designing effective real-time systems, a major influence on the development of the field has been the cost of computers. The 1960's saw the emergence of 'mini-computers', which made it financially feasible to dedicate individual computers to particular projects. Many real-time systems were set up at Harwell during this period and a Study Group on Computing and Data Processing reviewed and co-ordinated proposals; we also contributed to a Ministry of Technology Small Computers Working Party, which attempted (unsuccessfully) to stimulate the production of a suitable British minicomputer.

The costs of the hardware fell while the power of the computers rose - a trend which is occurring even more dramatically now with the availability of microprocessors and microcomputers(3).

This change in the economics of real-time systems has overturned the balance of project costs. In the 1960's it was difficult and expensive to make the hardware, as everything from sensors, actuators, datalinks, computers, to operator consoles and displays, had to be specially designed for the particular project; programming was considered a relatively inexpensive part of the work. By contrast, we now find it easy and relatively inexpensive to buy most of the equipment, as the components are available from a variety of sources, and to use compatible interfaces, but it takes a substantial time and effort to develop the programs.

3. Software Engineering Influences

Techniques for system design and software engineering have developed throughout the 1960's and 1970's, influencing programming methods, project management and programming languages. Although difficulties with real-time projects were the prime motivation for the work, the take-up of new techniques was relatively slow.

The general issues addressed by software engineering are those concerned with (i) the operational requirements of the systems to be produced, (ii) the achievement of these in practice and (iii) keeping the project manageable and the resulting implementation maintainable. The most important development has been the idea of high level programming languages, in which the program is expressed in an abstract form appropriate to the problem. Until recently, high-level programming languages have not dealt with the problems of real-time computing. Software engineering techniques in the late 1960's and early 1970's gave useful insights on systems design, emphasising the need for a clear specification and the importance of top-down as opposed to bottom-up design. However, the techniques seemed applicable only to small systems, and the particular real-time aspects were not considered. In developing much of the software we had to extend the ideas considerably and to introduce the principle of stratified design, where top-down principles were applied separately in a number of layers of facilities.

In the early days of real-time computing, the general feeling was that the specific issues were so dominant that conventional software engineering methods (as far as they were known then) could not be applied. We wrote the front-end and HUW programs in assembly language, as did all other real-time programmers, and then discovered how difficult it was to construct, test and maintain the resulting systems. However, it eventually became apparent that, in spite of the direct connection between computer and equipment, the greater part of the program deals with internal data structures and algorithms and so can be expressed in a high-level language. It is now felt that over 80% of a real-time program is actually sequential, so if software engineering methods are used to design this part, the rest can be given the special treatment it needs.

4. Particular Problems of Real-Time Programming

Because the computers in real-time systems are directly connected to special equipment, the design of the program must take account of the particular properties of that equipment. These establish constraints for the program which are quite unlike those of conventional computing and cannot be satisfactorily accommodated in a top-down analysis. The nature of the constraints has become clearer during the last ten years and we can now distinguish separate problems concerning timing, reliability and communication.

The interaction between the computer and its environment includes communication with other computers and with the operators who control the plant by means of the computer. The connections are through a number of input/output devices; the computer is required to respond to a variety of stimuli from the plant, within particular time intervals and with assurance of continuity over considerable periods of time. The particular input and output devices may not have been designed by the computer manufacturer, and it is likely that software written previously for the computer has no knowledge of them. The techniques for man-machine interaction have been developed arbitrarily, and experience on the human factors involved leaves too much to the intuition of the designer.

In the following sections we describe these problems in a little more detail, and outline the present position on their solutions.

4.1 Timing

The computer receives an input signal from the plant (a stimulus) and calculates a response which it must then send back to the plant within a given critical time. In the early systems it required careful programming to meet the time-constraint (hence the phrase 'real-time'), but the increased speed of computers has essentially eliminated this problem in its rudimentary form, but leaving a more subtle one.

In practice, a real-time system has many inputs and many outputs, between which there are many stimulus-response relations with their critical times. While the computer could easily satisfy any of these individually, there is substantial difficulty in satisfying them all collectively, particularly when they use inter-related data. A key development was the design of computers with interrupts, and the recognition of 'multi-tasking' as a necessary software technique. The idea of multi-tasking had arisen in the 1960's and this had a central role in the design of HUW. The essential idea was to write the program as a number of pieces each of which is executed sequentially, but with the interactions between them controlled by an executive or operating system(4). Multi-tasking does not necessarily satisfy the time constraints, but provides the necessary program structure so that critical parts can be distinguished and given sufficient priority(5).

4.2 Continuity of service

A computer in a real-time system is likely to be less expensive than the plant it controls; consequently, it is more important to keep the plant going than to get work through the computer. Characteristically, the plant (and perhaps the computer system) will include a certain amount of redundancy to cover the possibilities of equipment faults, which the computer may have to administer.

Faults may arise in the plant, which the computer can detect and deal with. Faults in the computer itself can also be detected, but it is not so easy to recover from them. Hardware faults can be treated by replicating and reconfiguring, but software faults are different. Faults in software arise not because of ageing but from unrecognised errors in design or assembly; if the software is replicated, each copy has the identical fault in it.

Early work gave attention to this problem. There were regular checks on the equipment being used and the computer hardware, with spares available to be reconfigured when necessary. Software faults were handled by using the same reconfiguration process, with a new copy of the software which acquires a reconstituted set of data. Any design errors in the software were not expected to be corrected as part of the immediate recovery process, but studied and improved as a software maintenance function. This illustrates the two principal approaches to software reliability(6): fault-intolerance, which means designing the software correctly and testing it adequately before use, and fault-tolerance, which means designing the software with internal checks and recovery during use, to catch any residual faults.

While the importance of program correctness has been appreciated generally, there is not so much awareness of techniques for fault-tolerant software design. Software redundancy requires alternative algorithms to achieve the same effect, with an automatic switch to another path if a fault is detected during the course of execution. No programming systems currently in use (before Ada) can support this mechanism, so it is very little used in practice.

4.3 Special input/output

The input/output devices used in real-time computing are frequently unknown to the designer of the computer or its operating system, and their particular properties are necessary for the proper instrumentation and control of the plant - they must not be hidden behind abstract interfaces where the programmer cannot refer to their key features. This is particularly true for time-related properties, when the device makes an interrupt request and the application program must respond to it promptly.

The solution usually adopted is to define a sub-program or macro (implemented in assembly language) for carrying out the particular functions on the device. Recent advances in programming languages (Modula(7) and Ada(2)) allow these basic actions to be programmed explicitly, which integrates the device handling with the rest of the program. The timing relations are handled in both Modula and Ada by treating interrupt handlers as tasks in a multi-tasking system.

Particular further problems arise with data links and man-machine interaction. Data links are special in that the information put into them might be lost en route and techniques are necessary to check correct receipt, perhaps with a noticeable time-delay between sending and acknowledgment. Work on data links at Harwell was begun in the late 1960's for connecting other small computers to the central mainframe through HUW, and suitable protocols were developed; however, further work on computer networks has shown that the underlying principles are not well understood, and generally agreed higher-level protocols are only now becoming settled.

Man-machine interaction is special for very different reasons. Early real-time systems used special keyboards, mimic-diagrams and displays to show operators the state of the system and allow them to state how it was to be controlled. Now there is much greater uniformity on the devices used (interactive terminals of various kinds) but the attention has switched to human engineering: how to enable theoperator to get a good view of the complex system, taking into account the needs to see different amounts of detail without loss of sensitivity to alarm conditions. It is being recognized that the real-time system can be as important for displaying the model of the plant to the operator as it is for controlling the plant in accordance with the operator's instructions.

5. Conclusion

The main impression I have of the progress in real-time computing during the last twenty years is that we have been building systems we did not really understand. Since the driving force in computing is to understand in order to be able to build more effective systems, this has given rise to feelings of unease and strong desires to discover what are the necessary underlying principles of the subject. To the extent that it is now possible to distinguish the major problems, we have made progress, but the possible solutions to those problems are by no means firmly settled yet.

The close connection betvween research ideas and their formulation in programs has meant that development follows rapidly in real-time computing, particularly in environments where there are strong interactions between scientists using computers in advanced ways and those investigating new techniques of meeting the requirements.

The developments during the last twenty-five years have been explosive, and the subject of real-time computing has emerged as a key topic for the practical application of computers. We are perhaps now in an adolescent stage, where there is much turmoil, searching for design principles, experimenting with new ideas, development of new components, and competition for standards. During the next quarter century, we might expect the rate of change to settle down as the principles become clearer and the necessary industrial standards in hardware, software and system design become established. The result will be cheaper and adequately reliable systems whose users will be only barely conscious that they involve computers at all.

1. H. Bisby, The CAMAC Standard. AERE-M.2507 (1972).

2. J.D. Ichbiah, Preliminary Reference Manual for the Programming Language Ada. SIGPLAN Notices, Vol. 14, No. 6, June 1979.

3. The Future of Real-Time Technology, Department of Industry, 1977. Ed. I. M. Barron.

4. E.W. Dijkstra, Co-operating Sequential Processes, in Programming Languages, Academic Press, 1968, Ed. F. Genuys.

5. N. Wirth, Towards a Discipline of Real-Time Programming, Comm. ACM 20, 577 (1977).

6. Infotech, System Reliability: A state-of-the-art report, 1978.

7. N. Wirth, Modula: A Language for Multiprogramming, in Software Practice and Experience, Vol. 7, No. 1 (Jan 1977) pp.3-35.

Epilogue

Reading this Special Progress Report brings back a flood of memories - of formative experiences, of friendships made, of the early days of atomic energy, of the inspiration of John Cockcroft - a host of things. I arrived in Harwell from Chalk River in 1946 and was assigned to the Nuclear Physics Division until one day I inadvertently applied a flame to some glass apparatus filled with hydrogen. I voluntarily transferred tó Theoretical Physics Division where I might do less damage. It was led by Klaus Fuchs, who had been doing even more damage. In the ensuing sequence of events I soon found myself in his seat as Head of the Division, two years before the formation of the Atomic Energy Authority. The manner in which the oldsters in the Division rallied behind their new and brash Head says much for their goodwill and tolerance.

The following years saw the broadening of the work of the Division into nuclear theory, solid state physics, atomic and plasma physics, all encouraged by the use of large-scale computing. The whole fascinating story is unfolded in the pages of this Report.

Very few of the original team now remain. Many made their reputations at Harwell and continued their work elsewhere. The interests have changed with the people and with the needs of the programme. But throughout its history the Division has been at the forefront in its chosen fields. It continues to bring high scholarship to bear upon the practical problems of the day. Its reputation is high. Long may it remain so.

The Lord Flowers, F.R.S.