RAJA RAMANNA

60TH BIRTHDAY FELICITATION VOLUME

Published by
Indian Academy of Sciences
in collaboration with
Indian National Science Academy
and
Indian Physics Association
RAJA RAMANNA

60TH BIRTHDAY FELICITATION VOLUME

Published by
Indian Academy of Sciences
in collaboration with
Indian National Science Academy
and
Indian Physics Association
© 1985 by the Indian Academy of Sciences
Dedicated to R. Ramanna
on his
60th Birthday
R. Ramanna (28 January 1925)
DEDICATION

Dr Raja Ramanna attained the age of 60 on 28 January 1985. He is well known to the scientific community in India and abroad for his dedicated contributions in various fields of scientific activity and especially to the atomic energy programme of the country. He has been closely connected with the Indian Academy of Sciences and had served on its Council for several years. He was also President of the Indian National Science Academy and the Indian Physics Association. These three organisations jointly publish Pramana (a journal of physics). It is therefore most appropriate that this special issue of the journal is brought out to commemorate the 60th birthday of Dr Ramanna.

Ramanna was born in Mysore and had his early education at Bangalore and later at the Madras Christian College, Madras. After obtaining his B.Sc. (Hons) degree in Physics from Madras Christian College in 1945, he went to Kings college, London and took his Ph.D. under Professor Allan May in 1949. He joined the Tata Institute of Fundamental Research the same year. In 1953 he was transferred to the Atomic Energy Establishment, Trombay (the present BARC) as the head of the Nuclear Physics Division. He continued to be a member of the TIFR Faculty. In 1962 he was appointed Director of the Physics Group in BARC and in 1972 he was appointed Director BARC and member for Research and Development in the Atomic Energy Commission. In 1973 he was appointed as part-time Chairman of Bharat Electronics Ltd., Bangalore. In 1978 he became Scientific Adviser to Defence Minister; Secretary to the Government of India, Defence Research; and Director General Defence Research and Development Organisation. In January 1981 he returned to BARC as Director and Secretary to Government of India in the Department of Atomic Energy and Member for Research and Development, Atomic Energy Commission. In 1983 he was appointed Chairman, Atomic Energy Commission.

Dr Ramanna has actively participated in the design of India's first reactor APSARA, and was in-charge of the utilisation of research reactors APSARA and CIRUS and the 5-5. MeV Van de Graaff Accelerator. He was the Chairman of the Variable Energy Cyclotron Committee constituted to set up a cyclotron at Calcutta and he was also the Chairman, Planning and Co-ordination Committee of the Reactor Research Centre at Kalpakkam for planning and implementing a fast reactor programme in India.

Ramanna has made outstanding contributions in several areas of nuclear science. This includes neutron thermalisation, reactor design, experimental and theoretical studies of low energy nuclear reactions with special reference to nuclear fission. His research work in the area of nuclear fission is particularly well-known. Under his leadership, pioneering work has been carried out in the studies pertaining to the emission of prompt neutrons,
gamma rays, K x-rays and light charged particles in fission. Extensive studies of the fragment mass, kinetic energy and angular distribution and the correlations between them both in the case of thermal and fast neutron induced fission have also been carried out. He has also made important contributions to the development of gas ionization cum scintillation detectors for the experimental research programmes in the areas of fission physics. In his theoretical research work, his outstanding contribution has been the development of the stochastic theory of fission based on the nucleon exchange mechanism. The importance of this mechanism in the heavy ion reactions also is now well recognised. Ramanna's role in other areas of peaceful utilization of nuclear energy are well known.

Another important contribution of Dr Ramanna is his strategy of organisation of science in developing countries. He has been Chairman of the 'NORA' Committee of the International Atomic Energy Agency, which looked after joint projects in Norway on Reactor Physics, and the India-Philippines-Agency Project where for the first time a successful co-operative effort of Asian scientists was organised. Currently he is a Member of the Scientific Advisory Committee to the Director-General, International Atomic Energy Agency.

As Member of the National Committee on Science and Technology he made available to the nation his deep understanding of the problems of development of science and technology. He is a member of the newly constituted Scientific Advisory Committee to the Cabinet.

Among the numerous honours he has received include the Shanti Swarup Bhatnagar Memorial Award for physical sciences 1963, Padma Shri in 1968, Padma Bhushan in 1973 and Padma Vibhushan in 1975. He was elected a Fellow of the Indian Academy of Sciences in 1966 and a Fellow of the Indian National Science Academy in 1973. He was awarded the Meghnad Saha Medal of the Indian National Science Academy for 1984. He was the President of the Physics Section of the Golden Jubilee Session of the Indian Science Congress. He played a leading role in the starting of the Indian Physics Association and supported it with great vigour. Dr Ramanna is welcomed in scientific gatherings not only for his technical expertise but also for the witty remarks and the charming manner with which he conducts the meetings. Those who are fortunate to know him well are aware of his skill in music, especially with piano.

On the occasion of Dr Ramanna attaining the age of 60 it is a matter of great pleasure to convey our best wishes on behalf of the Indian Academy of Sciences, the Indian National Science Academy and the Indian Physics Association and on behalf of the scientific community for many years of creative scientific activity and happy life.

Bangalore
February 1985

E. S. Raja Gopal
Chairman, Editorial Board

Pramāṇa
## CONTENTS

### Quantum Physics

<table>
<thead>
<tr>
<th>Topic</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantum supersymmetric generalisation of Bogomolnyi bounds</td>
<td>R Rajaraman</td>
<td>1</td>
</tr>
<tr>
<td>Special topics in the quantum theory of angular momentum</td>
<td>K Srinivasa Rao</td>
<td>15</td>
</tr>
<tr>
<td>Modified perturbation series for the anharmonic oscillator</td>
<td>Nazakat Ullah</td>
<td>27</td>
</tr>
<tr>
<td>On the Feigenbaum-Cvitanovic equation in the theory of chaotic</td>
<td>Virendra Singh</td>
<td>31</td>
</tr>
<tr>
<td>behaviour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantum chaos and fluctuation properties of regular and irregular</td>
<td>Jitendra C Parikh and Vijay B</td>
<td>39</td>
</tr>
<tr>
<td>spectra</td>
<td>Sheorey</td>
<td></td>
</tr>
<tr>
<td>Self-energy effect in a relativistic bound state problem</td>
<td>S N Biswas and Ashok Goyal</td>
<td>47</td>
</tr>
</tbody>
</table>

### Particle Physics

<table>
<thead>
<tr>
<th>Topic</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angular momentum and isospin properties of chiral fields</td>
<td>Klaus Goeke and J N Urbano</td>
<td>53</td>
</tr>
<tr>
<td>Relativistic harmonic confinement of quarks and gluons</td>
<td>S B Khadkikar</td>
<td>63</td>
</tr>
<tr>
<td>Isobaric degrees of freedom in nuclei as determined from nonrelativistic</td>
<td>C S Warke</td>
<td>69</td>
</tr>
<tr>
<td>quark model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The fifth interaction (universal long range force between spins)</td>
<td>T Pradhan, R P Malik and P C</td>
<td>77</td>
</tr>
<tr>
<td>Naik</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bethe-Salpeter dynamics for two-photon processes</td>
<td>A N Mitra, N Nimai Singh and R</td>
<td>95</td>
</tr>
<tr>
<td>Ramanathan</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Nuclear Physics

<table>
<thead>
<tr>
<th>Topic</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adiabatic time-dependent Hartree-fock theory with the consistency</td>
<td>A K Mukherjee and M K Pal</td>
<td>109</td>
</tr>
<tr>
<td>condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleon transport processes in fission and heavy ion reactions</td>
<td>S S Kapoor and V S Ramamurthy</td>
<td>119</td>
</tr>
<tr>
<td>Angular distribution in ternary fission</td>
<td>M M Sharma and G K Mehta</td>
<td>131</td>
</tr>
</tbody>
</table>

(Continued)
Radiochemical studies on fission of actinides

M V Ramaniah

Studies of light charged particle emission in fission at Trombay

S S Kapoor and D M Nadkarni

Development of thermal reactor lattice physics methods

B P Rastogi and H C Huria

Special features of the safety and control systems of the Dhruva reactor

A K Ray and S N Seshadri

Physics and engineering aspects of fast reactor safety

C V Sundaram and S M Lee

Engineering development and safety back-up for nuclear power programme

P R Dastidar and A Kakodkar

Accelerator development in India

A S Divatia and C Ambasankaran

Phenomenology of the Pokaran PNE experiment

R Chidambaram, S K Sikka and Satish C Gupta

### Atomic and Molecular Physics

Perspectives in neutron physics research

P K Iyengar and K Subba Rao

Risk evaluation and protection against ionizing radiation

V K Jain

Electron energy loss spectroscopy (EELS) of organic molecules in vapour phase: Design and fabrication of an indigenous EEL spectrometer

M S Hegde, V Jayaram, P Vishnu Kamath and C N R Rao

### Solid State Physics and Materials Science

Neutron diffraction studies of transition metal nitrides

N S Satya Murthy, S K Paranjpe, Marsongkohadi, Kurniadi Sumaamidjaja and M R L N Murthy

Melting in two dimensions—the current status

G Venkataraman and Sahoo

Defect structure studies using positron annihilation spectroscopy

R Krishnan and D D Upadhyaya

Radiation–induced structural changes in alloys

K Krishan and C V Sundaram

X-ray photoelectron spectroscopy for surface film analysis in corrosion research

P K Chauhan, H S Gadiyar and R Krishnan

Irradiation behaviour of nuclear fuels

P R Roy and D N Sah
Quantum supersymmetric generalisation of Bogomolnyi bounds

R RAJARAMAN
Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560012, India
Present address: Centre for Theoretical Physics, Massachusetts Institute of Technology,
Cambridge, Mass, USA

Abstract. We review here classical Bogomolnyi bounds, and their generalisation to
supersymmetric quantum field theories by Witten and Olive. We also summarise some recent
work by several people on whether such bounds are saturated in the quantised theory.

Keywords. Supersymmetry; gauge theories; quantum solitons; monopoles.

PACS No. 03 70

1. Introduction

Bogomolnyi published, in 1976, a very interesting paper (Bogomolnyi 1976) on some
aspects of classical topological soliton solutions of several relativistic field theories. His
procedure was generic in nature and covered such examples of solitons* as Kinks in two
(space-time) dimensions, vortices and flux lines in 3 dimensions as well as the famous
monopole solution in 4 dimensions. The thrust of his results was two-fold. Firstly he
showed that the energy of any finite-energy field configuration in these systems was
bounded from below by its topological index, apart from a known numerical factor.
Secondly, on requiring that this bound is saturated, one attains first order partial
differential equations—far simpler to solve than the parent field equations which are
second order non-linear partial differential equations. The familiar solitons like the
Kink, the flux lines and the monopole can be obtained as solutions of these first-order
equations. (It must be mentioned that the basic trick used by Bogomolnyi had been
employed slightly earlier, in the context of instantons, by Belavin et al 1975).

Bogomolnyi’s work was at the classical level. Meanwhile, methods have been
developed during the past decade for “quantising” solitons, to yield quantum soliton
particles in the corresponding quantum field theories. A detailed discussion of the
quantisation of solitons is given in the book by the present author (Rajaraman 1982).
One finds that these quantum solitons possess, in a suitably generalised sense, many of
the remarkable properties that the classical solitons did. One can ask therefore whether

* Strictly speaking, most of these solutions should be called solitary waves, rather than solitons. But here
we follow the practice in most of the recent literature of using the term solitons to also cover solitary waves, ie
localised, non-dispersive solutions. Note however, that contrary to the incorrect feeling that still persists in
some circles, solitary waves are not limited to (1 + 1) dimensions. Especially after the entry of particle
physicists in this field, many solitary wave solutions have been found in higher dimensions as well. Thus,
interesting gauge-theoretic solutions such as vortices in (2 + 1) dimensions and the ’tHooft-Polyakov
monopole in (3 + 1) dimensions are legitimate solitary waves.
Bogomolnyi-type bounds can also be obtained for quantum solitons, relating their exact quantum masses to their topological quantum numbers, and whether these bounds are saturated.

In an elegant and compact paper, Witten and Olive (1978) showed that for quantum supersymmetric extensions of some of the models considered by Bogomolnyi, an exact bound can be obtained for the masses of the quantum soliton states. In fact, Witten and Olive pointed out that the existence of the soliton sector affects the very algebra of supersymmetry (susy), introducing a central charge in the algebra in that sector. This central charge is intimately related to the topological charge of the soliton and, after some manipulations, yields a lower bound to the masses of all the states in the soliton sector. Having obtained the bounds, Witten and Olive also offered some speculations on whether quantum solitons actually saturate these bounds, as in fact they do at the classical approximation.

The question of whether these bounds are saturated at the quantum level is an important one. For, if they are, they would give us some handle on the exact quantum mechanical masses of these soliton particles. One must remember that calculating the mass-spectrum is one of the primary goals of particle physics. Yet hitherto no one has been able to calculate exactly, the mass of any massive bound state in any non-trivial 4-dimensional field theory! We emphasize that these quantum bounds apply not just to solitons in two space-time dimensions like the Kink or the sine-Gordon soliton, but also to some solitons in realistic four-dimensions as well, such as the (supersymmetric) monopole.

Recent investigations by several workers have pursued further the question of whether the Witten-Olive bounds are saturated and the related questions of whether the soliton mass and its topological index each receive quantum corrections in these susy models. The detailed results, as we will see, vary with the models.

In this article, we review the developments mentioned above. This is strictly a review—prepared specifically for the special volume in honour of Dr Raja Ramanna. There are no new results in this article, but only a synthesis of results already contained in the references cited. Also, in order to keep the length of this article within its prescribed bounds, we have to assume that the reader is familiar with the overall background in which our specialised topic is imbedded, which includes the basics of gauge theory, supersymmetry, classical and quantum solitons, and the 't Hooft-Polyakov monopole.

### 2. Classical Bogomolnyi bounds

Bogomolnyi inequalities at the classical level are well-known. Our chief concern here is their quantum generalisation. Nevertheless, for the sake of completeness, let us quickly sketch the classical bounds and their derivation. We will use as illustrations (1 + 1)-dimensional scalar field theories as well as the (3 + 1) dimensional Georgi-Glashow model which yields that famous soliton, the 't Hooft-Polyakov monopole (Georgi and Glashow 1972; 't Hooft 1974; Polyakov 1974).

Let us begin with the simple problem of a real scalar field theory in (1 + 1) dimensions, with any potential which has degenerate minima so that topological solitons can arise. The Lagrangian is

\[
L = \int dx \left[ \frac{1}{2} (\partial_{\mu} \phi) (\partial^{\mu} \phi) - \frac{1}{2} S^2 (\phi) \right],
\]
where without loss of generality, we can take the potential to be positive and to vanish at its absolute minima. We write this potential in the form \( \frac{1}{2} S^2(\phi) \) for the later convenience when we supersymmetrise the system in §3.

The classical energy functional associated with this Lagrangian is

\[
E = \int_0 dx \left[ \frac{1}{2} \left( \frac{\partial \phi}{\partial t} \right)^2 + \frac{1}{2} \left( \frac{\partial \phi}{\partial x} \right)^2 + \frac{1}{2} S^2(\phi) \right].
\]

Clearly, all finite energy configurations must approach one of the zeroes of \( S(\phi) \) at the spatial extremities \( x = -\infty \) and \( x = +\infty \). We are interested in a lower bound to the energy, and therefore need to consider only static configurations, where the (positive) kinetic energy is absent. The energy of static (time independent) fields \( \phi(x) \) is

\[
E = \int_{-\infty}^{\infty} dx \left[ \frac{1}{2} \left( \frac{\partial \phi}{\partial x} \right)^2 + \frac{1}{2} S^2(\phi) \right] = \int_{-\infty}^{\infty} dx \frac{1}{2} \left( \frac{\partial \phi}{\partial x} \right)^2 + \int_{-\infty}^{\infty} dx S(\phi) \frac{d\phi}{dx} \geq \left| \int_{-\infty}^{\infty} dx S(\phi) \frac{d\phi}{dx} \right| = 2 \left[ D(\phi(x)) \right]_{x=-\infty}^{x=\infty},
\]

where

\[
D(\phi) = \int d\phi S(\phi).
\]

This is the Bogomolnyi bound for this system and sets a lower bound on the energy of any field configuration. The bound is saturated, i.e. the equality in (3) is obeyed provided (i) \( \phi \) is static, and (ii) it obeys

\[
d\phi/dx = \pm S(\phi)
\]

Meanwhile, the field equation arising from (1), for static configurations, is

\[
\frac{d^2 \phi}{dx^2} = S(\phi) \frac{dS}{d\phi}
\]

The first integral of this field equation is just the saturation condition (5). Hence any static classical solution, including the topological soliton, will satisfy (5) and therefore saturate the bound (3). Thus we know, without explicitly finding the soliton solution, that its energy will be just \( D(\phi_2) - D(\phi_1) \) where \( \phi_2, \phi_1 \) are the boundary values of the soliton at \( x = \pm \infty \) respectively. Clearly this bound, depending as it does only on boundary values of the field, is a topological index of the soliton.

As an illustration, consider the kink solution of the double-well potential, where

\[
S(\phi) = \sqrt{\frac{\lambda}{2}} \left( \phi^2 - \frac{\mu^2}{\lambda} \right).
\]

Then

\[
D(\phi) = \sqrt{\frac{\lambda}{2}} \left( \frac{\phi^3}{3} - \frac{\mu^2}{\lambda} \phi \right).
\]

The kink solution goes from \( \phi_1 = -\mu/\sqrt{\lambda} \) at \( x = -\infty \) to \( \phi_2 = +\mu/\sqrt{\lambda} \) at \( x = \infty \). Hence the bound is
Meanwhile, on explicitly solving the field equation, the kink solution has the well-known form

$$\phi_{\text{kink}} = \frac{\mu}{\sqrt{\lambda}} \tanh \frac{\mu x}{\sqrt{2}}$$

with a classical energy (the classical kink mass) equal to

$$M_{\text{kink}}^{(0)} = \frac{2\sqrt{2} \mu^3}{3\lambda}.$$  \hspace{1cm} (8)

This exactly equals the bound (7), as expected.

As a second example, let us turn to the Georgi-Glashow model, which is a non-abelian \((3 + 1)\) dimensional gauge theory, with a triplet of scalar fields \(\phi^a\) coupled to a triplet of \(SU(2)\) gauge fields \(A^a\). The Lagrangian density is

$$\mathcal{L} = \frac{1}{4} (D_\mu \phi^a)(D^\mu \phi^a) - \frac{\lambda}{4} (\phi^a \phi^a - C^2)^2 - \frac{1}{4} F_{\mu \nu}^a F^{a \mu \nu}$$  \hspace{1cm} (9)

with \(\mu, \nu = 0, 1, 2, 3; a = 1, 2, 3, (D_\mu \phi^a) \equiv \partial_\mu \phi^a + g \epsilon_{abc} A_\mu^b \phi^c\)

and

$$F_{\mu \nu}^a \equiv \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g \epsilon_{abc} A_\mu^b A_\nu^c.$$  \hspace{1cm} (10)

Although this system is more complicated than (1), Bogomolnyi showed that a bound can be obtained in a similar way. Consider again the energy functional for static configurations, in the \(A_0^a = 0\) gauge.

$$E = \int d^3 x \left[ \frac{1}{4} F_{ij}^a F_{ij}^a + \frac{1}{2} D_1 \phi^a D_1 \phi^a + \frac{\lambda}{4} (\phi^a \phi^a - C^2)^2 \right]$$  \hspace{1cm} (11)

with \(i, j = 1, 2, 3\). Finiteness of energy requires, as \(|x| \to \infty\),

$$\phi^a \phi^a \to C^2,$$  \hspace{1cm} (12a)

and

$$F_{ij}^a, D_i \phi^a \to 0.$$  \hspace{1cm} (12b)

Each field configuration permitted by the boundary condition (12a) corresponds to a mapping of a two-sphere \(S_2\) into another \(S_2\). Since the second homotopy group \(\pi_2(S_2) = \mathbb{Z}\), such configurations fall into homotopy classes characterised by an integer-valued index, given by

$$N = \frac{1}{8\pi} \int_{S_2} d\sigma_1 \epsilon_{ijk} \epsilon^{abc} (\hat{\phi}^a) (\hat{\phi}^b) (\hat{\phi}^c),$$  \hspace{1cm} (13)

where \(\hat{\phi}^a = \phi^a/|\phi| = \phi^a/C\). Recall also the result that a configuration with homotopy index \(N\) carries magnetic charge \(m = N/g\). For a detailed discussion of these well known results, see Rajaraman (1982). The energy in (11) can be written, after a little algebra, in the form

$$E = \frac{4\pi C}{g} N + \int d^3 x \left[ \frac{1}{4} (F_{ij}^a - \epsilon_{ijk} D_k \phi^a)^2 + \frac{\lambda}{4} (\phi^a \phi^a - C^2)^2 \right]$$

$$\geq \frac{4\pi C}{g} N = 4\pi C m$$  \hspace{1cm} (14)
Quantum supersymmetric generalisation of Bogomolnyi bounds

where \( m \) is the magnetic charge of the configuration. This is Bogomolnyi's inequality for the Georgi-Glashow model. In the so-called Prasad-Sommerfield limit (Prasad and Sommerfield 1975), where \( \lambda \to 0 \) with \( C \) fixed, the bound (14) is saturated, provided

\[
F_{ij}^a = \varepsilon_{ijk} (D_k \phi)^a. \tag{15}
\]

A solution of (15), by virtue of minimising the static energy in any given \( N \)-sector, will also be a classical solution of the field equations. In fact, the well-known Prasad-Sommerfield solution for the monopole

\[
\phi^a = C r^a \left[ \coth (rgC) - \frac{1}{rgC} \right]; \quad A_0^a = 0; \quad \text{and}
\]

\[
A_i^a = \frac{1}{gr} \varepsilon_{aib} r^b \left( 1 - \frac{rgC}{\sinh rgC} \right)
\]

satisfies (15) and lies in the \( N = 1 \) sector. Correspondingly, its classical mass,

\[
M_{\text{mono}}^{(0)} = 4\pi C/g, \tag{16b}
\]

saturates the bound (14). Note that (15) is a first-order equation, like (7), although the parent field equations derived from (10) would be of second order.

Similar results can also be obtained for the Nielsen-Olesen vortex lines, the \( CP_n \) and non-linear \( O_3 \) models in \( (2 + 1) \) dimensions. Note that the discussion in this section, based on Bogomolnyi's (1976) paper, has been entirely at the classical level.

3. The Witten-Olive bounds

Witten and Olive (1978) considered the quantised and supersymmetric extension of the models discussed in the previous section. First take the two-dimensional example and consider the Lagrangian density

\[
\mathcal{L} = \frac{1}{2} \left[ (\partial_\mu \phi)^2 - S^2(\phi) + \bar{\psi} (i \gamma_\mu \gamma^a - S'(\phi)) \psi \right]; \tag{17}
\]

where \( \psi \) is a Majorana spinor in two dimensions. The two \( \gamma \)-matrices could be taken as, say \( \gamma^0 = \sigma_2 \) and \( \gamma^1 = -i \sigma_1 \). \( S(\phi) \) is any function of the real scalar field \( \phi \) which permits topological solitons. It must have degenerate minima. This Lagrangian is supersymmetric under the transformation

\[
\phi \to \phi + \tilde{\xi} \psi; \quad \psi \to \psi - ( - i \gamma_\mu \partial^\mu \phi - S(\phi)) \tilde{\xi}; \quad \bar{\psi} \to \bar{\psi} + \bar{\xi} (i \gamma_\mu \partial^\mu \phi - S) \tag{18}
\]

where \( \tilde{\xi}, \bar{\xi} \) are Grassmann numbers. For \( S = (\lambda/2)^{1/2} (\phi^2 - C^2) \), the system (17) is the supersymmetric extension of the double well system (1).

The supercharge which generates the transformations (18) is also a Majorana spinor \( Q \) given by

\[
Q = \int dx (\gamma^a \partial_\mu \phi + i S) \gamma^0 \psi. \tag{19}
\]

In our representation of \( \gamma \)-matrices, charge conjugation is just complex conjugation. Hence the Majorana spinors \( Q \) and \( \psi \) have only real components:

\[
Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}; \quad \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}; \tag{20}
\]
with \[ Q_1 = \int dx \left[ (\partial_0 \phi + \partial_1 \phi) \psi_1 - S \psi_2 \right], \quad (21) \]
\[ Q_2 = \int dx \left[ (\partial_0 \phi - \partial_1 \phi) \psi_2 + S \psi_1 \right]. \]

Using the canonical commutation rules for the fields \( \phi \) and \( \psi \) it is straightforward to check that
\[ (Q_1)^2 = H + P, \]
\[ (Q_2)^2 = H - P, \]
and
\[ \{Q_1, Q_2\} = 2 \int dx S(\phi) \frac{d\phi}{dx} = 2 [D(\phi)]_{\phi(x=\infty)} - \frac{\phi(x=-\infty)}{2}, \]
where \( H \) and \( P \) are the total energy and momentum operators respectively, for the system (17), and \( D(\phi) \equiv \int d\phi S(\phi) \). Equations (22a, b) can be recognised as part of the "usual" susy algebra in 2 dimensions. But in the "usual" algebra, the anticommutator in (22c) would have been taken to vanish, and indeed it does vanish in the vacuum sector, where \( \phi(x) \) takes the same value at \( x = -\infty \) and \( x = +\infty \), and so will \( D(\phi(x)) \).

But, in the soliton sector, where \( \phi(\infty) \neq \phi(-\infty) \), \( D(\phi(\infty)) \) need not equal \( D(\phi(-\infty)) \). To see this, note that a static topological soliton will go from some \( \phi = \phi_1 \) at \( x = -\infty \) to some other \( \phi = \phi_2 \) at \( x = +\infty \), where \( \phi_1 \) and \( \phi_2 \) are two neighbouring distinct zeroes of \( S(\phi) \). Therefore \( \phi_1 \) and \( \phi_2 \) will also be two neighbouring extremes of \( D(\phi) \), one of them a minimum and the other the next maximum. Clearly \( D(\phi_1) \) cannot equal \( D(\phi_2) \). Thus, in the sector of states based on the topological soliton, the operator \( T \) is non-vanishing, and forms (as can be verified) a central charge in the susy algebra. (See, however, remarks by Schonfeld (1979) on the validity of (22) in the face of boundary conditions.)

The quantum version of the bound (3) follows immediately from the algebra (22). We have
\[ H = \frac{1}{2} (Q_1^2 + Q_2^2) = \frac{1}{2} [(Q_1 \pm Q_2)^2 \mp \{Q_1, Q_2\} + ] \]
\[ = \frac{1}{2} [(Q_1 \pm Q_2)^2 \mp T] \geq \frac{1}{2} |T|. \]

This operator inequality is the quantum generalisation of (3), obtained by Witten and Olive for the susy system (17). The last step in (23) holds because \((Q_1 \pm Q_2)^2\) is a real non-negative operator. Equation (23) is an operator inequality, i.e. it holds for the expectation values of both sides taken between any state of the quantum field theory associated with (17). In the vacuum sector \( \langle T \rangle = 0 \) for all states, so that (23) is trivial. But in the soliton sector it is not. The quantum soliton-particle at rest is the lowest energy state in this sector. Let us denote it by \(|\text{sol}\rangle \). Then the full quantum mass of the soliton, \( M_\text{sol} \), obeys
\[ M_\text{sol} = \langle \text{sol} | H | \text{sol} \rangle \geq \frac{1}{2} |\langle \text{sol} | T | \text{sol} \rangle| . \]

(24)
Note that (24) is still in the form of an inequality. We will return later to the question of whether the equality holds.

A similar bound can be obtained for the quantum supersymmetric monopole by considering a susy extension of the Georgi-Glashow model (9–10). The Lagrangian, in (3 + 1) dimensions, is

\[
\mathcal{L} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} + \frac{1}{2} i \bar{\psi}_j \gamma_\mu (D_\mu \psi_j) + \frac{1}{2} (D_\mu \phi_j)^a (D^a_\mu \phi_j) + \frac{g^2}{2} \text{Tr} \left\{ \left[ \phi_1, \phi_2 \right] \right\}^2 + \frac{i g}{2} \epsilon_{jk} \text{Tr} \left\{ \left[ \psi^l, \psi^k \right] \phi_1 + \left[ \bar{\psi}^l, \gamma_5 \psi^k \right] \phi_2 \right\}
\]

where \( \psi_j^a, j = 1, 2; a = 1, 2, 3 \) form a pair of isotriplet Majorana spinors, \( \phi_i^a \) is a scalar and \( \phi_2^a \) a pseudoscalar isotriplet, and \( F^a_{\mu\nu} \) the gauge field tensor of the \( \text{su}(2) \) gauge group. We have also used the popular notation of representing isotriplets \( \phi_i^a \) and \( \phi_j^a \) by matrix valued fields \( \phi_j \equiv (i a/2) \phi_j^a \) and \( \psi_j = (i a/2) \psi_j^a \) respectively in the last two trace terms in (25). The model (25) enjoys \( N = 2 \) supersymmetry whose generators are

\[
Q_i = \text{Tr} \int d^3x \left\{ \sigma^{\mu\nu} F_{\mu\nu} \phi_i + \epsilon_{ij} \left( D_5 \phi_1 \right) + \left( D_5 \phi_2 \right) \gamma_5 \psi^i \right\} + g \left\{ \phi_1, \phi_2 \right\} \gamma^0 \gamma_5 \psi_i,
\]

where \( i = 1, 2 \) and the spinorial indices of \( Q \) and \( \psi \) are suppressed. When the anticommutators of the \( Q_i \) are evaluated using the canonical commutation rules of the fields, one obtains

\[
\{ Q_i, Q_j \} = \delta_{ij} \gamma^a P_a + \epsilon_{ij} \left( U + \gamma_5 V \right),
\]

where

\[
U \equiv \int d^3x \partial_i (\phi_1^a F^a_{oi} + \frac{1}{2} \epsilon_{ijk} \phi_2^a F^a_{jk}),
\]

\[
V \equiv \int d^3x \partial_i (\phi_2^a F^a_{oi} + \frac{1}{2} \epsilon_{ijk} \phi_1^a F^a_{jk}).
\]

Note that the boson potential in (25) has the positive form \( \text{Tr} \left\{ \left[ \phi_1, \phi_2 \right] \right\}^2 \). When \( \phi_1 = \phi_2 = 0 \), this vanishes. Correspondingly, one choice of the vacuum is the symmetric one, with \( \left\langle \phi_1 \right\rangle_{\text{vac}} = \left\langle \phi_2 \right\rangle_{\text{vac}} = 0 \). However, \( \text{Tr} \left\{ \left[ \phi_1, \phi_2 \right] \right\}^2 \) also vanishes if the matrices \( \phi_1 \) and \( \phi_2 \) commute, ie if the iso-vectors \( \phi_1^a \) and \( \phi_2^a \) are parallel in internal space, regardless of the magnitudes \( |\phi_1^a| \) and \( |\phi_2^a| \). This choice \( \left\langle \phi_1^a \right\rangle_{\text{vac}} \propto \left\langle \phi_2^a \right\rangle_{\text{vac}} \), with either non-zero, corresponds to spontaneous breaking of the \( \text{su}(2) \) gauge symmetry. Now, consider the charges \( U \) and \( V \) in (28). Both are integrals of divergences and can be written as surface integrals at infinity. In the symmetric case, \( \left\langle \phi_1^a \right\rangle_{\text{vac}} = \left\langle \phi_2^a \right\rangle_{\text{vac}} = 0 \), both \( \phi_1^a \) and \( \phi_2^a \) will vanish at spatial infinity for any physical state, and hence \( U = V = 0 \). Then (27) reduces to the “usual” susy algebra. It is only when \( \left\langle \phi_1^a \right\rangle_{\text{vac}} \) or \( \left\langle \phi_2^a \right\rangle_{\text{vac}} \) is non-zero that the central charges \( U \) and \( V \) come into play.

The quantum Bogomolnyi-type bound follows easily from (27). A simple trick, used by Witten and Olive is to exploit the fact that the system (25) is also chirally invariant. Under a chiral rotation, the fields \( \phi_1^a \) and \( \phi_2^a \) rotate into one another, and therefore so do \( U \) and \( V \). Perform a rotation such that \( \left\langle V \right\rangle = 0 \). Take the expectation value of (27) between any energy-momentum eigenstate (whether solitonic or otherwise) at rest.
Then (27) gives

\[ \langle \{ Q_i, Q_j \} \rangle > = \delta_{ij} M + \epsilon_{ij} \gamma^0 \langle U \rangle, \]

(29)

where \( M \) is the invariant mass of the state. Note that \( Q_i \) (or \( Q_j \)) is also a spinor in Dirac space, with spinor index \( \alpha \) (or \( \beta \)) which we had suppressed so far, in addition to the index \( i \) (or \( j \)) which takes values 1, 2. Considered as an \( 8 \times 8 \) matrix, \( \langle \{ Q_{ia}, Q_{j\beta} \} \rangle > \) is a positive matrix. It will have real non-negative eigenvalues. Therefore so will the matrix \( \delta_{ij} M + \epsilon_{ij} \gamma^0 \langle U \rangle \), thanks to (29). But \( \epsilon_{ij} \gamma^0 \langle U \rangle \) clearly has eigenvalues \( \pm \langle U \rangle \). Hence,

\[ M \geq |\langle U \rangle|. \]

(30)

This result was obtained after setting \( \langle V \rangle = 0 \) through a chiral rotation. The general result is clearly a chirally invariant generalisation of (30), viz.

\[ M^2 \geq \langle U \rangle^2 + \langle V \rangle^2. \]

(31)

This is the Witten-Olive bound for the susy Georgi-Glashow model, for the mass of any state, in terms of the expectation values of \( U \) and \( V \) in that state. For, the case \( \langle \phi^a_1 \rangle \text{vac} = \langle \phi^a_2 \rangle \text{vac} = 0 \), which corresponds to full unbroken \( su(2) \) gauge symmetry, \( \langle U \rangle = \langle V \rangle = 0 \) for any state. Then the bound (31) is trivial. But, for all symmetry-broken cases, with either \( \langle \phi^a_1 \rangle \text{vac} \) or \( \langle \phi^a_2 \rangle \text{vac} \) non-zero, the bound (31) is non-trivial. In such cases, it is applicable not only in the solitonic sectors, but also in the vacuum sector.

That is, (31) can be used to get a lower bound on the exact quantum masses of the monopoles and dyons, as well as the vector bosons and the Higgs bosons of this susy theory. Thus, this result of Witten and Olive is a truly remarkable one.

4. Saturation of the bounds

The quantum Witten-Olive bounds (23) and (31) reduce, in the appropriate classical limit, to the Bogomolnyi bounds discussed in §2, and are, furthermore, saturated classically by soliton solutions. For the \((1 + 1)\) dimensional scalar field case, it is obvious that the quantum bound (23), where the operator \( T \) is given in (22c), reduces to just the Bogomolnyi bound (3) in the classical limit. This classical bound, as we showed in §2, is saturated by the soliton. For the \((3 + 1)\) dimensional gauge theory, the relationship of the classical bound (14) to the quantum bound (31) may be made more transparent as follows. Using chiral symmetry freedom, we can consider the case where only \( \langle \phi^a_1 \rangle \text{vac} \neq 0 \), with \( \langle \phi^a_2 \rangle \text{vac} = 0 \). Let the modulus \( |\langle \phi^a_1 \rangle \text{vac}| = C \), with the internal space direction of \( \langle \phi^a_1 \rangle \text{vac} \) arbitrary. For all finite energy systems, topological or otherwise, \( |\phi^a_1| \) must then tend to \( C \) and \( \phi^a_2 \) must tend to zero as \( |x| \to \infty \). The gauge group \( su(2) \) is broken down to \( U(1) \) by such a vacuum. The associated electromagnetic field \( F_{\mu \nu}^a \) has the gauge invariant form (t Hooft 1974)

\[ F_{\mu \nu}^\text{em} = \frac{1}{C} \phi^a \mathcal{F}_{\mu \nu}^a - \frac{1}{g C^3} \epsilon_{abc} \phi^b_1 (D_\mu \phi_1)^b (D_\nu \phi_1)^c \]

\[ \to \frac{1}{C} \phi^a \mathcal{F}_{\mu \nu}^a \]

(32)

since finiteness of energy requires that \( (D_\mu \phi_1)^a \to 0 \) faster than \( 1/|x|^{3/2} \) as \( |x| \to \infty \).

Now, consider the charges \( U \) and \( V \) defined in (28). Since \( \phi^a_2 \to 0 \) as \( x \to \infty \), we have,
using Gauss' theorem,
\begin{equation}
U = \int_S d\sigma_i (\phi^a_i F_{ai}^a) = C \int_S d\sigma_i F_{ai}^c m = 4\pi C q,
\end{equation}
\begin{equation}
V = \int_S d\sigma_i (\phi^a_i F_{ai}^a) = C \int_S d\sigma_i F_{ai}^c m = 4\pi C m,
\end{equation}
where \( q \) and \( m \) are just the total electric and magnetic charge operators. Hence the bound (31) can be written as
\begin{equation}
M^2 \geq (4\pi C)^2 [\langle q \rangle^2 + \langle m \rangle^2],
\end{equation}
where \( \langle q \rangle \) and \( \langle m \rangle \) are the expectation values of the total electric and magnetic charges of the state in question. Thus this bound applies non-trivially to particles which carry either electric or magnetic charge, or both. Equation (34) is still the same quantum bound as (31), but rewritten in terms of more familiar charges. In the classical limit, when applied to static topological configurations in the \( A_0 = 0 \) gauge (which carry zero electric charge) it reduces to precisely the classical Bogomolnyi bound (14). As pointed out in §2, the exact single-monopole solution, in the Prasad-Sommerfield limit, saturates this classical bound. For states which carry no magnetic charge (these will be non-solitonic states, in the \( N = 0 \) homotopy sector) the bound in (34) is still useful, giving
\begin{equation}
M = 4\pi C |\langle q \rangle|.
\end{equation}
The familiar "quanta" of the fields, such as the Higgs boson, the massive Vector (\( W \)) boson and the photon of this theory come in this category. Notice again that at the classical ("tree") level, the \( W \) boson has electric charge \( q/4\pi \) and mass \( qC \), thus saturating the bound classically. The photon is a special case. Thanks to unbroken \( U(1) \) gauge invariance, the photon (\( q = 0 \)) has exactly zero-mass quantum theoretically, and will infact saturate the exact quantum bound (35). For the other particles, \( \text{ie} \) the monopole, the \( W \)-boson, and the Higgs boson of this model, as well as for the solitons of the two-dimensional model, while precise quantum bounds exist in the form of (24) and (34), their saturation at the quantum level calls for further discussion. The rest of this section is devoted to this question.

Let us begin with solitons of \((1+1)\) dimensional scalar field theories, whose quantum bound is given in (24). There is no rigorous closed result available, as far as we know, establishing that the quantum soliton must saturate the bound. What most people have attempted is to evaluate separately the quantum corrections to both sides of the bound, \( \text{ie} \) to the soliton mass and to its topological index \( T \). Then one can see if these corrections equal one another. Of course this is not a completely definitive way of answering the question. Quantum corrections are generally calculated using a semi-classical loop expansion, in powers of \( \hbar \). In practice such a calculation can be done only up to some given order, usually to order \( \hbar \). If the two sides of (24) agree upto \( O(\hbar) \), that by itself does not guarantee that they will agree to all orders in \( \hbar \). (See however an ingenious and indirect argument by Witten and Olive, based on counting of states, suggesting that the monopole does exactly satisfy the quantum bound. See also Imbimbo and Mukhi (1984a) for the 2-dimensional soliton). Of course, if the two sides of (24) do not equal each other upto order \( \hbar \), then we can be sure that the bound is not saturated.

Let us first consider calculations of the \( O(\hbar) \) corrections to the soliton mass. The general principles of soliton quantisation, particularly in \((1+1)\) dimensions, are by now
well known (for reviews, see Rajaraman 1975; Jackiw 1977; Coleman 1977) so that we need not present the details of this calculation here. But there are some special features which arises in susy models that one should be careful about. The soliton’s quantum mass to $O(h)$ is given by

$$M_{\text{sol}}^{(1)} = M_{\text{sol}}^{(0)} + \frac{1}{2} \hbar \left( \sum_{w_n > 0} w_n \right) - \frac{1}{2} \hbar \left( \sum_{e_n > 0} e_n \right) + M_{\text{c1}}, \quad (36)$$

where $w_n^2$ and $e_n$ are respectively the eigenvalues of the boson fluctuation equation and the fermion Dirac equation in the background of the soliton. $M_{\text{c1}}$ is the contribution of the $O(h)$ renormalisation counter term. For the system (17), the eigenvalue equations determining $w_n^2$ and $e_n$ are

$$\left( - \frac{d^2}{dx^2} + (S')^2 + SS'' \right) \eta_n(x) = w_n^2 \eta_n(x) \quad (37)$$

$$\left( - i \alpha \frac{d}{dx} + \beta S' \right) \psi_n(x) = e_n \psi_n(x) \quad (38)$$

where $S' = \left[ \frac{d}{d\phi} S(\phi) \right]_{\phi_{\text{sol}}(x)}$ and $S'' = \left[ \frac{d^2 S}{d\phi^2} \right]_{\phi_{\text{sol}}(x)}$. D’Adda and Di Vecchia (1978) made the following important observation regarding (37) and (38). Let us write the two-component spinor $\psi_n$ as

$$\begin{pmatrix} \psi_n^{(+)} \\ \psi_n^{(-)} \end{pmatrix}.$$  

It will be convenient to use the representation $\beta = \sigma_2$, $\alpha = -\sigma_1$. Then, up on squaring the Dirac Hamiltonian in (38), we get

$$(i \sigma_1 d_x + \sigma_2 S')^2 \psi_n = \left( - d_x^2 + (S')^2 - \sigma_3 \frac{d}{dx} S' \right) \psi_n = e_n^2 \psi_n. \quad (39)$$

Since the soliton obeys $d\phi_{\text{sol}}/dx = -S(\phi_{\text{sol}})$, $dS'/dx = (S'')d\phi_{\text{sol}}/dx = -S''S$, (39) yields, for each component of $\psi_n$,

$$(- d_x^2 + (S')^2 + SS'') \psi_n^{(+)} = e_n^2 \psi_n^{(+)} \quad (40a)$$

$$(- d_x^2 + (S')^2 - SS'') \psi_n^{(-)} = e_n^2 \psi_n^{(-)}. \quad (40b)$$

One can see that thanks to supersymmetry the upper component of the spinor, $\psi_n^{(+)}$, obeys the same eigenvalue equation as (37), obeyed by the boson fluctuations $\eta_n$. Thus any solution of (37) would also serve as a solution of (40a) with the same eigenvalue and vice versa. The associated $\psi_n^{(-)}$ could then be determined from the Dirac equation (38) i.e

$$\psi_n^{(-)} = \frac{1}{e_n} (id_x + iS') \psi_n^{(+)}. \quad (41)$$

As evident from (41), this matching between boson and fermion fluctuations may not hold for zero-energy ($e_n = 0$) modes, but we need not worry about these for our purposes, since they make no contribution to the energy in (36).

Upto this point, these observations of D’Adda and Di Vecchia are interesting and correct. However, from this they concluded that $\Sigma w_n - \Sigma e_n = 0$ and that the soliton receives no quantum corrections to its mass upto order ($h$). This is not correct for two
Quantum supersymmetric generalisation of Bogomolnyi bounds

reasons, as pointed out by Schonfeld (1979) and by Kaul and Rajaraman (1983):
(i) The counter-term contribution \( M_{ct} \) has been neglected. Normal ordering counter terms proportional to \( SS" \) exist for these models consistent with supersymmetry and make a non-zero (and in fact divergent) contribution \( M_{ct} \).
(ii) A more subtle point is that \( (\Sigma w_n - \Sigma e_n) \) does not vanish even though, loosely speaking, every non-zero eigenvalue \( w \) of (37) also occurs as an eigenvalue of (38) and vice versa. The reason is that, apart from a few discrete levels, the spectra of both (37) and (38) are continuous, and the densities of continuum level in the two cases are different. The simplest way to understand this difference in density states is to put the system in a box of length \( L \), and later take \( L \to \infty \). Then the Boson eigenfunction \( \eta_n(x) \) obeys boundary conditions
\[
\eta_n(-L/2) = \eta_n(L/2),
\] (42a)
and
\[
\frac{d\eta_n}{dx}(-L/2) = \frac{d\eta_n}{dx}(L/2),
\] (42b)
as appropriate to the second order differential operator in (37). However, the fermion eigenfunction \( \psi_n(x) \) obeys the first order Dirac equation (38), and correspondingly, obeys first order boundary conditions, but, for both components:
\[
\psi_n^{(+)}(-L/2) = \psi_n^{(+)}(L/2),
\] (43a)
and
\[
\psi_n^{(-)}(-L/2) = \psi_n^{(-)}(L/2).
\] (43b)
The condition (43b) can be reduced, using (41), to a condition on \( \psi_n^{(+)} \):
\[
[(d_x + S')\psi_n^{(+)}(x)]_{-L/2} = [(d_x + S')\psi_n^{(+)}(x)]_{L/2}.
\] (43c)
We can see that although both \( \eta_n(x) \) and \( \psi_n^{(+)}(x) \) obey the same differential equations (37) and (40a), they must satisfy different boundary conditions. The condition (42a) is the same as (43a), but (42b) is not generally the same as (43c), especially for topological solitons for which \( S' = (dS/d\phi)\phi(x) \) is not the same at \( x = \pm L/2 \), as \( L \to \infty \). The eigenvalues of differential operators are specified both by the differential form of the operator as well as the boundary conditions. Therefore, the set of boson eigenvalues \( w_n^2 \) and fermion eigenvalues \( e_n^2 \) will in fact be different, for any finite \( L \), however large. As \( L \to \infty \), both spectra will merge into the same continuum, but with different spectral densities. The difference \( \Sigma w_n - \Sigma e_n \) will not vanish in general.

Once we are aware of these pitfalls, the evaluation of the fermion and boson densities of states \( \rho_f(E) \) and \( \rho_B(E) \) using the correct boundary conditions (42–43), is straightforward. One finds that the fermion level density is in fact the average of the densities of the two equations (40a) and (40b), as one would expect from symmetry between the upper and lower components of the spinor \( \psi_n \). Given the densities of states, the fluctuation energy \( (\Sigma w_n - \Sigma e_n) = \int E dE (\rho_B(E) - \rho_f(E)) \) can be calculated, added on to the counter term to obtain the quantum soliton mass (see Schonfeld 1979; Kaul and Rajaraman 1983; their results have been rederived, using more elegant methods and without recourse to boundary conditions, by Imbimbo and Mukhi 1984a).
To illustrate the results, consider again the example of the Kink of the SUSY double-well problem. For this system
\[
S = \left( \frac{\lambda}{2} \right)^{1/2} \left( \phi^2 - \mu^2/\lambda \right),
\] (44a)
while the counter term added to $S$ is

$$S_{\text{ct.}} = -\hbar \left( \frac{\lambda}{2} \right)^{1/2} \left[ \int_0^\infty \frac{dk}{2\pi} \frac{1}{(k^2 + 2\mu^2)^{1/2}} + K \right],$$

(44b)

where $K$ is an arbitrary finite constant. On computing $\frac{1}{2} \Sigma w_n - \frac{1}{2} \Sigma e_n + S_{\text{ct.}}$, one finds, for the $O(h)$ quantum Kink mass,

$$M_{\text{kink}}^{(1)} = m_1^3/3\lambda + \frac{h}{\sqrt{12\pi}} m_1,$$

(45)

where we have eliminated dependences on $\mu$ and $K$ by using $m_1$, which is the renormalised mass, to $O(h)$, of the boson in that model. In the classical limit, the second $O(h)$ term in (45) is absent and $m_1$ can be replaced by the “tree” level boson mass $m_0 = (2)^{1/2}\mu$ in the double-well problem. Then (45) reduces to the classical kink mass $M_{\text{kink}}^{(0)}$ given in (8). However there is not much meaning in calculating the difference between $M_{\text{kink}}^{(1)}$ and $M_{\text{kink}}^{(0)}$ to see “how much” quantum correction the kink-mass acquires. In quantum field theory, thanks to ultraviolet divergences and their removal by subtraction schemes which are not unique, the relation between the same physical quantity evaluated to different orders in $h$, contains arbitrariness. For instance, the one-loop renormalised boson mass $m_1$ can be evaluated in terms of the “bare” boson mass $m_0 = (2)^{1/2}\mu$ by standard perturbation techniques, but the relation between $m_1$ and $m_0$ is arbitrary up to a constant. Any attempt to write $m_1$ in terms of $m_0$ and insert it into (45) so that it may be compared with the classical result (8), will be fraught with ambiguity (see Kaul and Rajaraman 1983 for more details). But (45) as it stands, is meaningful. It gives the one-loop quantum mass of the kink in terms of one-loop boson mass. One can also meaningfully compare (45) with $\langle \text{sol} \mid T \mid \text{sol} \rangle$, also evaluated to one-loop level, to see if the bound (24) is saturated.

The one-loop correction to $\frac{1}{2} \langle \text{sol} \mid T \mid \text{sol} \rangle$ has been evaluated independently by Imbimbo and Mukhi (1984a) as well as by Chatterjee and Majumdar (1984). These one loop corrections come from two sources:

(i) Recall that $\frac{1}{2}T = [D(\phi(x))]_{x=0}^\infty$. Classically, we evaluated this by inserting the classical function $\phi = \phi_{\text{sol}}(x)$. In quantum theory $\phi(x, t)$ is an operator, which we write in terms of the shifted field operator $\eta(x, t)$,

$$\phi(x, t) = \phi_{\text{sol}}(x) + \eta(x, t).$$

Then, at each point $(x, t)$,

$$\langle \text{sol} \mid D(\phi) \mid \text{sol} \rangle = D(\phi_{\text{sol}}) + (dD/d\phi)\phi_{\text{sol}} \langle \text{sol} \mid \eta \mid \text{sol} \rangle$$

$$+ \frac{1}{2} (d^2D/d\phi^2)\phi_{\text{sol}} \langle \text{sol} \mid \eta^2 \mid \text{sol} \rangle + \ldots$$

(46)

Since $T$ involves $D(\phi)$ only at $x = \pm \infty$, note that for the kink system (44)

$$D(\phi_{\text{sol}}) = \mp \frac{(2)^{1/2}\mu^3}{3\lambda}; \quad \frac{dD}{d\phi}(\phi_{\text{sol}}) = 0; \quad \text{and}$$

$$\frac{d^2D}{d\phi^2}(\phi_{\text{sol}}) = \pm (2)^{1/2}, \quad \text{as } x \to \pm \infty.$$ 

(47)

Hence

$$\langle \text{sol} \mid D(\phi(x = \pm \infty, t)) \mid \text{sol} \rangle$$

$$= \pm \frac{(2)^{1/2}\mu^3}{3\lambda} \pm \frac{1}{2}(2)^{1/2}\mu \langle \text{sol} \mid \eta^2(\pm \infty, t) \mid \text{sol} \rangle$$

(48)
Quantum supersymmetric generalisation of Bogomolnyi bounds

The two-point function \( \langle \text{sol}|\eta^2(x, t)|\text{sol} \rangle \) can be easily obtained, to leading order, by using the expansion in terms of the normal modes \( \eta_n(x) \) in (37).

\[
\eta(x, t) = \left[ \sum_n \frac{a_n}{(2\omega_n)^{1/2}} e^{-i\omega_n t}.\eta_n(x) + \text{h.c.} \right]
\]

with \( [a_n, a_m^+] = \hbar \delta_{n,m} \).

This yields

\[
\langle \text{sol}|\eta^2(\pm \infty, t)|\text{sol} \rangle = \hbar \int_0^{\infty} \frac{dk}{2\pi} \frac{1}{(k^2 + 2\mu^2)^{1/2}}
\]

(50)

The two-point function will be ultraviolet divergent, but this is cancelled by the counter term (44b) which leads to

\[
D_{\text{c.l.}}(\phi_{\text{sol}}(x = \pm \infty)) = \left( \int d\phi S_{\text{c.l.}}(\phi) \right) \phi_{\text{sol}}(\pm \infty)
\]

\[
= \mp \frac{\hbar \mu}{(2)^{1/2}} \left[ \int_0^{\infty} \frac{dk}{2\pi} \frac{1}{(k^2 + 2\mu^2)^{1/2}} + K \right].
\]

(51)

On adding (50) and (51), one gets a finite result, which when cast in terms of the renormalised boson mass \( m_1 \) yields

\[
\frac{1}{2} \left| \langle \text{sol}|T|\text{sol} \rangle \right| = \frac{m_1^2}{3\lambda} + \frac{\hbar m_1}{(12)^{1/2}\pi} + O(h^2)
\]

(52)

in agreement with the quantum soliton mass in (45).

Thus we see that for a \((1 + 1)\)-dimensional scalar field soliton like the kink, the Witten-Olive bound is saturated at the \( O(h) \) quantum level. In fact, a stronger result appears in the recent work by Yamagishi (1984), where the Witten-Olive inequality as well as its saturation in \( O(h) \) quantum theory are worked out at the level of densities—ie relating the expectation values of the Hamiltonian density and the density of the topological index \( T \). That the bound holds to \( O(h) \) makes it plausible that it may hold to all orders in \( h \). An argument that it indeed does so hold, has been briefly advanced in the Imbimbo-Mukhi work.

Turning to the \((3 + 1)\) dimensional susy extension of the Georgi-Glashow model (25) an evaluation of the quantum corrections to the susy monopole’s mass has recently been done by Kaul (1984). Although the underlying principles and pitfalls are similar to those in \((1 + 1)D\) models, the actual calculation is more difficult. One has to compute fluctuation — energies associated with the two species of isotriplet Fermi fields \( \psi_1^a \) and \( \psi_2^a \), the two species of scalar fields \( \phi_1^a \) and \( \phi_2^a \), the gauge field \( A_\mu^a \), and ghost fields, and that too in 3 space-dimensions. All this has been ably done by Kaul (1984). He finds that, once again, the boson and fermion fluctuation energies do not cancel, and leave behind an ultraviolet-divergent residue. This divergence, again, is cancelled by counter terms. Further, the resulting answer for the one-loop monopole mass has the same form as the classical mass (16b) provided one replaces the vacuum-expectation value \( C \) and the coupling constant \( g \) by their renormalised values. That is

\[
M^{(1)}_{\text{mono}} = 4\pi \left( \frac{C}{g} \right)_{\text{ren}}.
\]

(53)
Turning to the right side of the quantum bound (34), as applied to the monopole, it will become, after one-loop corrections, \(4\pi(Cm)_{\text{ren}}\). Assuming that as required by the exact Dirac condition, \(m_{\text{ren}} = 1/g_{\text{ren}}\), the bound is saturated.

In the last stages of his proof of (53), Kaul uses the relation \(g_{\text{ren}}C_{\text{ren}} = gC\), ie that this product \(gC\) receives no renormalisation correction in this model. The validity of this assumption as a gauge—and renormalisation-scheme-independent statement has been questioned in a very recent preprint by Imbimbo and Mukhi (1984b). This preprint reached us just when we were completing this review, and we have not had the opportunity to digest its contents. We will merely report that these authors also study the quantum mass corrections for the \(N = 2\) susy monopole (as well as the \(N = 4\) susy extension), by using trace theorems which are a generalisation of what they had employed in their earlier paper quoted above. Their differences with Kaul notwithstanding they too find that the quantum bound is saturated for the \(N = 2\) susy monopole.

References

Chatterjee A and Majumdar P 1984 Preprint, Saha Inst. Nucl. Phys. TNP 84/5
Coleman S 1977 in New phenomena in sub-nuclear physics (ed.) A Zichichi (New York: Plenum)
Jackiw W 1977 Rev. Mod. Phys. 49 681
Imbimbo C and Mukhi S 1984a LPTENS Preprint 84/04
Imbimbo C and Mukhi S 1984b TIFR Preprint T H 84-23
Kaul R K 1984 CERN Preprint Th. 3840
Polyakov A M 1974 JETP Lett. 20 194
Rajaraman R 1982 Solitons and instantons (Amsterdam: North Holland)
Yamagishi H 1984 MIT Preprint CTP 1158
Special topics in the quantum theory of angular momentum

K SRINIVASA RAO
The Institute of Mathematical Sciences, Madras 600 113, India

Abstract. Two special topics in the quantum theory of angular momentum are discussed in this article. They are: (i) the relationship between the coupling and recoupling coefficients (for two and three angular momenta, respectively) and sets of generalized hyper-geometric functions of unit argument; and (ii) the 'non-trivial' or polynomial zeros of angular momentum coefficients and their classification.

Keywords. Quantum theory; angular momentum; coupling coefficients; recoupling coefficients; hypergeometric functions.

PACS No. 02.90; 03.65; 21.90

1. Introduction

The new developments in the quantum theory of angular momentum, during the past 25 years following the dramatic discovery of new symmetry properties of the angular momentum coupling and recoupling coefficients (for two and three angular momenta, respectively) by Regge (1958, 1959) led to the subject itself becoming a part of the Encyclopedia of Mathematics and its Applications. Volumes 8 and 9 of this series, by Biedenharn and Louck (1981) are entitled, “Angular momentum in quantum physics” and “The Racah-Wigner algebra in quantum theory”, respectively. In Chapter 5, of Vol. 9, twelve topics are developed and these establish the diverse relationships between concepts in angular momentum theory and other areas of mathematics. The new developments, in the words of Smorodinskii and Shelepin (1972) are “more and more intertwined with various sections of algebra, multidimensional geometry, topology, projective geometry, analytic function theory, the theory of special functions, differential equations, combinatorial analysis and the calculus of finite differences. One could say that the theory of Clebsch-Gordan coefficients takes on the character of a new kind of calculus, going far beyond the scope of the classical theory”.

In this article, we devote our attention to only two special topics of recent origin, where we have made some significant contributions. They are: (i) the relationship between the coupling (Clebsch-Gordan) and recoupling (Racah) coefficients and sets of generalized hypergeometric functions of unit argument, and (ii) the 'non-trivial' or polynomial zeros of the Clebsch-Gordan (3-\(j\)) and Racah (6-\(j\)) coefficients and their classification.
2. The relationship between the Clebsch-Gordan and Racah coefficients and generalized hypergeometric functions

2.1 The Clebsch-Gordan coefficient

Here we show that the series which is given in literature (see for instance, Biedenharn and Louck 1981) for the Clebsch-Gordan or 3-j coefficient belongs to a set of six representations and that there exist, correspondingly, a set of six \(_3F_2(1)s\), which is necessary and sufficient to account for the 72 symmetries of the 3-j coefficient.

The 3-j coefficient is defined by:

\[
\begin{pmatrix}
  j_1 & j_2 & j_3 \\
  m_1 & m_2 & m_3
\end{pmatrix} = \delta(m_1 + m_2 + m_3)(-1)^{j_1 + j_2 - m_3}\Delta(j_1, j_2, j_3)
\]

\[
\times \prod_{i=1}^{3} [(j_i + m_i)! (j_i - m_i)!]^{1/2}
\]

\[
\times \sum (1)(-1)^t \left[ t! \prod_{k=1}^{2} (t - \alpha_k)! \prod_{l=1}^{3} (\beta_l - t)! \right]^{-1},
\]

where

\[
\max(\alpha_1, \alpha_2) \leq t \leq \min(\beta_1, \beta_2, \beta_3),
\]

\[
\beta_1 = j_1 - m_1, \beta_2 = j_2 + m_2, \beta_3 = j_1 + j_2 - j_3,
\]

\[
\alpha_1 = j_1 - j_3 + m_2 = j_1 - m_1 - (j_3 + m_3),
\]

\[
\alpha_2 = j_2 - j_3 - m_1 = j_2 + m_2 - (j_3 - m_3),
\]

and \(\Delta(xyz) = [(x + y - z)! (x - y + z)! (-x + y + z)! / (x + y + z + 1)!]^{1/2}\). Equation (1) is invariant to permutations of the three angular momenta (column permutations) and to space reflections \((m_i \rightarrow -m_i)\), thereby exhibiting a 12-element symmetry group. Regge (1958) arranged the nine integer parameters referred to by Racah (1942), into a 3 x 3 square symbol and represented the 3-j coefficient as:

\[
\begin{pmatrix}
  j_1 & j_2 & j_3 \\
  m_1 & m_2 & m_3
\end{pmatrix} = \begin{array}{ccc}
  -j_1 + j_2 + j_3 & j_1 - j_2 + j_3 & j_1 + j_2 - j_3 \\
  j_1 + m_1 & j_2 - m_2 & j_3 - m_3 \\
  j_1 + m_1 & j_2 + m_2 & j_3 + m_3
\end{array} = \| R_{ik} \|.
\]

The fact that all sums of columns and rows are equal to \(J = j_1 + j_2 + j_3\) leads to the following nine relations amongst the elements of

\[
R_{lp} + R_{mp} = R_{nq} + R_{nr},
\]

for cyclic permutations of both \((lmn) = (123)\) and \((pqr) = (123)\). Regge (1958) discovered that there exists a 72-element symmetry group, comprising the aforementioned 12-element symmetry group, for the 3-j coefficient, due to the invariance of \(\| R_{ik} \|\) to its column and row permutations and a reflection about its diagonal.

Racah (1942) pointed out that substitution for the summation index \(t\), the argument of one of the five factorials in (1) only leads to some symmetry properties of the angular momentum coupling coefficient. It is straightforward to show that the five substitution procedures lead to five series representations. We show here that these together with (1) constitute a set of six series representations for the 3-j coefficient, which is necessary and sufficient to account for its 72 symmetries. Since the 72-element symmetry group is evident when the 3-j coefficient is represented by \(\| R_{ik} \|\), it is worthwhile to note that the
Quantum theory of angular momentum

set of six series representations can be written in the following compact notation:

\[ R_{ik} = \delta(m_1 + m_2 + m_3) \prod_{i,k=1}^{3} \left[ R_{ik} !/(J + 1) \right]^{1/2} \]
\[ \times (-1)^{\theta(pqr)} \sum_s \left[ s! (R_{2p} - s)! (R_{3q} - s)! (R_{1r} - s)! \right] \]
\[ \times (s + R_{3r} - R_{2p})! (s + R_{2r} - R_{3q})! \]^{-1} \quad (4)

for all permutations of \((pqr) = (123)\), with

\[ \theta(pqr) = \begin{cases} R_{3p} - R_{2q} & \text{for even permutations of (123)} \\ R_{3p} - R_{2q} + J & \text{for odd permutations of (123)} \end{cases} \]

Let the six series representations be denoted by the Roman numerals (I) to (VI), corresponding to the six permutations of \((pqr)\) in (4): (123), (231), (312), (132), (321) and (213), respectively. This set of six series representations can also be obtained by permuting the indices (123) in (1), and remembering that the series acquires an additional phase factor of \((-1)^{4j}\), for odd permutations. In establishing the one-to-one correspondence between the series obtained by the substitution procedures and that obtained by permuting the indices in (1), use is made of the fact that \(4j_i\) is an even integer so that \((-1)^{4j_i}\) is always positive. The following observations can now be made, with the help of (3) where necessary:

(a) the column permutations of \(R_{ik}\) are in one-to-one correspondence with the six series representations, thereby spanning the whole set given by (4);

(b) the permutation of the second and third rows of \(R_{ik}\) which corresponds to space reflection, exchanges (I) \(\leftrightarrow\) (VI), (II) \(\leftrightarrow\) (V) and (III) \(\leftrightarrow\) (IV);

(c) the permutation of the first and second rows of \(R_{ik}\) exchanges (I) \(\leftrightarrow\) (V), (II) \(\leftrightarrow\) (IV) and (III) \(\leftrightarrow\) (VI);

(d) the permutation of the first and third rows of \(R_{ik}\) exchanges (I) \(\leftrightarrow\) (IV), (II) \(\leftrightarrow\) (VI) and (III) \(\leftrightarrow\) (V);

(e) the cyclic permutation of rows, (123) \(\rightarrow\) (231), of \(R_{ik}\) permutes anti-cyclically the first three and the last three of the six series representations, amongst themselves, as: (I) \(\rightarrow\) (III) \(\rightarrow\) (II) \(\rightarrow\) (I) and (IV) \(\rightarrow\) (VI) \(\rightarrow\) (V) \(\rightarrow\) (IV);

(f) the cyclic permutation of rows, (123) \(\rightarrow\) (312), of \(R_{ik}\) permutes cyclically the first three and the last three of the six series representations, amongst themselves, as: (I) \(\rightarrow\) (II) \(\rightarrow\) (III) \(\rightarrow\) (I) and (IV) \(\rightarrow\) (V) \(\rightarrow\) (VI) \(\rightarrow\) (IV);

(g) the reflection of \(R_{ik}\) about its diagonal interchanges (I) \(\leftrightarrow\) (III) leaving the other four series representations invariant;

(h) the reflection of \(R_{ik}\) about its skew-diagonal interchanges (IV) \(\leftrightarrow\) (V) leaving the others invariant;

(i) each series representation is left invariant to the combined operation of an odd column permutation and the space reflection; as well as to either the Regge symmetries (Appendix 1) as such or Regge symmetries on which are superposed an even column permutation or an odd column permutation and the space reflection;
(j) each series representation exhibits 12 of the 72 distinctly different symmetries of the 3-j coefficient and this 12-element symmetry group is isomorphic to the product of permutation groups of 3 objects ($R_{2p}$, $R_{3q}$, $R_{1r}$) and 2 objects ($R_{3r} - R_{2p}$, $R_{2r} - R_{3q}$).

From the observations (a) to (j) it follows that the set of six series representations is necessary and sufficient to account for the 72 known symmetries of the 3-j coefficient.

Conventional ways (refer, for instance, deShalit and Talmi 1963) of establishing the symmetry relations are tedious and some (Rose 1955) necessarily resort to the substitution procedures. Our observation that $\|R_{ik}\|$ is in fact a set of six series representations enables one to establish the symmetries easily. Notice that the first three series corresponding to even column permutations are separated from the last three series which correspond to odd column permutations, in the sense that the latter contain the additional phase factor, $(-1)^j$.

The series representations can be rearranged into generalized hypergeometric functions (Slater 1966) of unit argument by using:

$$n! = \Gamma (n + 1)$$

for the factorials and whenever the summation index in the argument of the $\Gamma$ function is negative resorting to:

$$\Gamma (n) \Gamma (1 - n) = \pi \csc \pi n$$

(6)

to replace it by a $\Gamma$-function with an argument containing a positive index of summation. This procedure of rearrangement yields in the case of the 3-j coefficient (Srinivasa Rao 1978; Venkatesh 1978) the following set of six $3F_2(1)$s:

$$\|R_{ik}\| = \delta (m_1 + m_2 + m_3) \left[ \prod_{i,k=1}^{3} R_{ik}! / (J + 1) \right]^{1/2}$$

$$\times (-1)^{\eta(pqr)} [\Gamma (1 - A, 1 - B, 1 - C, D, E)]^{-1}$$

$$\times 3F_2 (ABC; DE; 1),$$

(7)

where

$$A = -R_{2p}, B = -R_{3q}, C = -R_{1r}, D = 1 + R_{3r} - R_{2p},$$

$$E = 1 + R_{2r} - R_{3q} \quad \text{and} \quad \Gamma (x, y, \ldots) = \Gamma (x) \Gamma (y) \ldots,$$

for all permutations of $(pqr) = (123)$. This set of six $3F_2(1)$s given here are all of the van der Waerden's form (Smorodinskii and Shelepin 1972)—the $3F_2(1)$ given by (5·21) of this reference corresponds to $(pqr) \equiv (123)$.

2.2 The Racah coefficient

Here we show that it is possible to obtain two equivalent sets of series representations for the Racah (6-j) coefficient, which can be rearranged into two equivalent sets of $4F_3(1)$s. These sets are shown to be related to each other through the property of reversal of series of the generalized Saalschutzian hypergeometric function.

The conventional expression for the Racah coefficient (Biedenharn and Louck 1981) is given by:

$$\begin{cases} abe \\ cde \end{cases} = (-1)^{a+b+c+d} W (abcd; ef)$$
where the range of $P$ is

$$P_{\text{min}} \leq P \leq P_{\text{max}},$$

with $P_{\text{min}} = \max(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$,

$$P_{\text{max}} = \min(\beta_1, \beta_2, \beta_3),$$

$$\alpha_1 = a + b + e, \quad \alpha_2 = c + d + e, \quad \alpha_3 = a + c + f, \quad \alpha_4 = b + d + f,$$

$$\beta_1 = a + b + c + d, \quad \beta_2 = a + d + e + f, \quad \beta_3 = b + c + e + f,$$

$$N = \Delta(abc)\Delta(cde)\Delta(acf)\Delta(bdf),$$

and $\Delta(xyz)$ defined as before (below equation (1)).

The 6-7 coefficient is invariant under:

(i) the $3! = 6$ column permutations and

(ii) the four interchanges of any two elements in the first row of the 6-7 symbol with the corresponding elements in the second row (this will be referred to as ‘row’ permutation in the text). These 24 symmetries constitute the classical tetrahedral symmetry group of the 6-7 coefficient.

In passing it is to be noted that this series form (8) can be found for the first time in Regge’s (1958) article. Racah (1942) dealt with only the series expansion obtained by substituting $s = \beta_1 - P$ in (8), which corresponds to one of a set of three representations, to be discussed later.

Once the series is in the form (8), it is easy to see that the 6-7 coefficient exhibits the 144-element symmetry group, due to its invariance under the permutation of the four $\alpha$’s and the three $\beta$’s. It was Regge (1958) who dramatically discovered six more symmetries (Appendix 2) and established that the Racah (6-7) coefficient exhibits 144 symmetries and not only the 24 tetrahedral symmetries.

By setting in (8), $s = \beta_k - P, k = 1, 2, 3$, in succession, we get the set I of three series expansions:

$$\left\{ \begin{array}{l} \alpha b e \\ d c f \end{array} \right\} = N (-1)^{\beta_k} \sum_s (-1)^s (\beta_k - s + 1)!$$

$$\times \left[ \prod_{i=1}^{4} (\beta_k - \alpha_i - s)! \prod_{j=1}^{3} (s + \beta_k - \beta_j)! \right]^{-1}.$$  \hspace{1cm} (9)

Notice that a series belonging to this set I exhibits only 48 of the 144 symmetries, due to the permutation of all the four $\alpha$’s but only two of three $\beta$’s, since $\beta_k$ is now in the numerator in (9).

Instead, if we set in (8), $s = P - \alpha_i, l = 1, 2, 3, 4$, in succession, we obtain (Srinivasa Rao and Venkatesh 1977) the following set II of four series representations:

$$\left\{ \begin{array}{l} \alpha b e \\ d c f \end{array} \right\} = N (-1)^{\alpha_i} \sum_s (-1)^s (\alpha_i + s + 1)!$$

$$\times \left[ \prod_{i=1}^{4} (s + \alpha_i - \alpha_i)! \prod_{j=1}^{3} (\beta_j - \alpha_i + s)! \right]^{-1}.$$  \hspace{1cm} (10)

where we notice that a series belonging to set II (10) exhibits only 36 of the 144
symmetries, arising due to the permutation of all the three $\beta$'s, but only three of the four $\alpha$'s since $\omega_1$ is now in the numerator in (10).

Thus, the 144 symmetries of the Racah (6-$j$) coefficient are exhibited by the single series expansion (8), or by the set I of three series representations given by (9), or, equivalently, by the set II of four series representations given by (10).

When (9) is rewritten in terms of generalized hypergeometric functions of unit argument, we get (Srinivasa Rao et al 1975), the set I of three generalized Saalschutzian hypergeometric functions of unit argument:

$$\left\{ \begin{array}{c}
abce \\
dcfa
\end{array} \right\} = (-1)^{E+1} \frac{\Gamma(1 - A, 1 - B, 1 - C, 1 - D; F, G)}{\Gamma(1 - A, 1 - B, 1 - C, 1 - D)}^{-1} \times {}_4F_3(ABCD; EFG; 1),$$

(11)

where

$$A = e - a - b, \quad B = e - c - d, \quad C = f - a - c, \quad D = f - b - d,$$

$$E = -a - b - c - d - 1, \quad F = e + f - a - d + 1, \quad G = e + f - b - c + 1$$

and $\Gamma(xyz \ldots) = \Gamma(x)\Gamma(y)\Gamma(z) \ldots$. Superposing the column permutations of $\{abce\}$ on the $\,_4F_3(1)$ in (11), yields us the set I of three $\,_4F_3(1)$s. We note that the superposition of a 'row' permutation of $\{abce\}$ on the $\,_4F_3(1)$ in (11) results only in a permutation of the numerator and denominator parameters amongst themselves in a given $\,_4F_3(1)$ belong to this set I.

When (10) is rearranged into a set of hypergeometric functions, we obtain (Srinivasa Rao and Venkatesh 1977) the following set II of four $\,_4F_3(1)$s for the 6-$j$ coefficient:

$$\left\{ \begin{array}{c}
abce \\
dcfa
\end{array} \right\} = (-1)^{A-1} \frac{\Gamma(1 - A', 1 - B', 1 - C', 1 - D'; E', F', G')}{{}_4F_3(1)}^{-1} \times {}_4F_3(A'B'C'D'; E'F'G'; 1),$$

(12)

where

$$A' = a + c + f + 2, \quad B' = c - d - e, \quad C' = a - b - e, \quad D' = f - b - d,$$

$$E' = a + c - b - d + 1, \quad F' = a + f - d - e + 1, \quad G' = c + f - b - e + 1.$$
Quantum theory of angular momentum

which is nothing but the $4F_3(1)$ belonging to set II given by (12) except for a trivial permutation amongst the numerator and denominator parameters of the generalized hypergeometric function.

Thus, we find that the set I and set II of $4F_3(1)$s for the 6-$j$ coefficient are not independent but are related to one another by the property of reversal of series. It is intriguing to note that due to reversal of series property the set I of three $4F_3(1)$s yields the set II of four $4F_3(1)$s and vice versa.

It has been possible for us (Srinivasa Rao and Venkatesh 1978; Srinivasa Rao 1981) to show the advantages of using the sets of $pF_q(1)$s in the numerical computation of the 3-$j$ and the 6-$j$ coefficients. The $pF_q(1)$s can be computed numerically using Horner's rule (Lee 1966) for polynomial evaluation as:

$$
pF_q(x_1 x_2 \ldots x_p; \beta_1 \beta_2 \ldots \beta_q; z) = \left[ 1 + \frac{x_0}{y_0} \left( z + \frac{x_1}{y_1} \left( z + \frac{x_2}{y_2} (z + \ldots) \right) \right) \right].
\tag{14}
$$

where

$$x_i = \prod_{j=1}^{p} (x_j + i) \quad \text{and} \quad y_i = (i + 1) \prod_{k=1}^{q} (\beta_k + i).$$

The use of (14) places our approach of numerical computation on the same footing as that of Wills (1971) and in the nested form the number of multiplications is minimum. In fact, the expression of Bretz (1976) is identical to the set II of $4F_3(1)$s given by (12). Thus, the set of $pF_q(1)$s which have been shown to be necessary and sufficient to account for the symmetries of the 3-$j$ and the 6-$j$ coefficient have been found to be not only useful but also time-saving when compared to the best available computer (Fortran) programs of Wills (1971) and Bretz (1976) by about 5 to 15%.

D'Adda et al (1974) and Raynal (1978, 1979) make use of Whipple's (1925, 1936) work on the symmetries of $3F_2$ functions, well-poised $7F_6$ functions and Saalschutzian $4F_3$ functions, all with unit argument, to study the 3-$j$ and the 6-$j$ symbols generalized to any arguments. Raynal (1978) generalized one of the formulas for the 3-$j$ symbol of SU (2) to obtain a generalized $3F_2(1)$ with complex parameters. Biedenharn and Louck (1981), discuss this aspect of the Racah-Wigner algebra—viz the relationship of the 3-$j$ and the 6-$j$ coefficients to generalized hypergeometric functions—as a special topic in Vol. 9 of the Encyclopedia in Mathematics and its Applications and summarize the results of our work presented here in a different notation.

3. ‘Non-trivial’ or polynomial zeros of angular momentum coefficients and their classification

Here we show that the polynomial or ‘non-trivial’ zeros of the Clebsch-Gordan (3-$j$) and the Racah (6-$j$) coefficients can be classified according to their degree. The majority of the zeros tabulated to-date, hitherto considered as non-trivial, are in fact polynomial zeros of degree one, arising due to the existence of binomial expansions for these coefficients.

3.1 Clebsch-Gordan coefficient

Trivial zeros of the 3-$j$ coefficient are those which arise due to either the triangular inequality not being satisfied by $j_1, j_2$ and $j_3$; or due to $J (= j_1 + j_2 + j_3)$ being odd when
$m_1 = m_2 = m_3 = 0$ (which can be easily verified as the consequence of the symmetry property of the 3-j coefficient under spatial reflection: $m_i \rightarrow -m_i$). The 'non-trivial' zeros are identified as the zeros of the polynomial part of the 3-j coefficient, defined in (4). They are also called as "structural" zeros.

Sato and Kaguei (1972), used the concept of generalized powers to obtain a formal binomial expansion for the 3-j coefficient. Following the procedure given by Sato (1955) and using the symbolic notation for generalized powers:

$$P(x) = P!/(P-x)!, \quad (15)$$

we can rearrange the set of six series representations given by (4) into the form:

$$\|R_{ik}\| = \delta(m_1 + m_2 + m_3) \prod_{i,k}^{3} \left[ R_{ik} \right]^{(J+1)} 1^{1/2} (-1)^{0(pqr)}
\times \left[ \Gamma(n+1, C_u + 1, C_v + 1, B_{rp} + n + 1, B_{rq} + n + 1) \right]^{-1}
\times \{ (B_{rp} + n) (B_{rq} + n) - C_u C_v \}^{(n)}, \quad (16)$$

where $n = \min(R_{2p}, R_{3q}, R_{1r})$, $C_u, C_v$ represent the $R_{ik}$'s in the triple $(R_{2p}, R_{3q}, R_{1r})$ other than the minimum, $B_{rp} = R_{3r} - R_{2p}$ and $B_{rq} = R_{2r} - R_{3q}$. The expression (16) is a generalization of the formal binomial expansion obtained by Sato and Kaguei (1972), for their result can be obtained by putting $(pqr) = (231)$ in (16), while (16) itself holds for all the six permutations of $(pqr) = (123)$.

Obviously, since the generalized power (15) is exact for $n = 1$, (since $P^{(1)} = P$) the binomial form for the 3-j coefficient explicitly reveals further structure zeros of this coefficient. These can be shown (Srinivasa Rao and Rajeswari 1984a) to be polynomial zeros of degree one. We can simply locate all the polynomial zeros of degree one of the 3-j coefficient, independent of the numerical values of the arguments of the 3-j coefficient, with the help of the simple factor:

$$(1 - \delta_{x,y} \delta_{n,1}), \quad (17)$$

where $x$ and $y$ are given by:

$$x = R_{mr} R_{kp} \quad \text{and} \quad y = R_{mp} R_{kr},$$

with $(lmk)$ and $(pqr)$ corresponding to specific permutations of (123) when $n = R_{pq}$.

Polynomial zeros of the 3-j coefficient were listed, for $J (= j_1 + j_2 + j_3) \leq 27$ by Varshalovich et al. (1975). Bowick (1976) used the Regge symmetries for the 3-j and the 6-j coefficients to obtain listings of only the Regge-inequivalent polynomial zeros. The number of terms in the polynomial part of the 3-j coefficient is governed by: $n = \min(R_{2p}, R_{3q}, R_{1r})$ and we use this simple prescription to classify the polynomial zeros tabulated by Bowick (1976) and by Biedenharn and Louck (1981) into those which correspond to degree 1, 2 and 4. The classified tables can be found in Srinivasa Rao and Rajeswari (1984b). From the tables, we conclude that for $J \leq 27$, of the 36 polynomial zeros, 21 are polynomial zeros of degree 1, revealed by the exact binomial expansion for this coefficient, represented by the multiplicative factor (17).

3.2 The Racah coefficient

Trivial zeros of the Racah coefficient are those which arise due to a violation of one of the four triangular inequalities to be satisfied by the six angular momentum arguments.
The existence of a class of zeros of the 6-j coefficient, not due to the aforesaid reason has been called as polynomial, ‘non-trivial’ or structure zeros by Koozekanani and Biedenharn (1974).

Sato (1955) expressed the Racah coefficient symbolically as a binomial expansion. Following his notation but making a different set of substitutions, corresponding to those required for getting the set I of series expansions given by (9), given below:

\[ s = \beta_0 - P, \quad 0 \leq s \leq n, \quad n = \beta_0 - \alpha_0, \]
\[ A_i = \beta_0 - \alpha_i \quad (i = p, q, r), \quad B_j = \beta_j - \beta_0 \quad (j = u, v), \]

where the indices \( p, q, r \) and \( u, v \) are used for those values of the \( \alpha \)'s and \( \beta \)'s other than \( P_{\text{min}} = \max (\alpha_1, \alpha_2, \alpha_3, \alpha_4) = \alpha_0 \) and \( P_{\text{max}} = \min (\beta_1, \beta_2, \beta_3) = \beta_0 \), we obtain (Srinivasa Rao and Venkatesh 1977) the binomial expansion:

\[
\binom{abc}{def} = N (-1)^{\beta_0} \left[ \Gamma(n+1, A_p+1, A_q+1, A_r+1, B_u+n+1, B_v+n+1) \right]^{-1} 
\times \left( (B_u+n)(B_v+n) - A_p A_q A_r (\beta_0+1)^{-1} \right)^{(n)}
\]

where we used the notation:

\[ P^{(x)} = P/(P-x)! \quad \text{and} \quad P^{(-x)} = 1/P^{(x)} \]

and regarded the form \( P^{(x)} \) to represent the generalized powers of \( P \). It follows that (19) is an exact binomial expansion for the power \( n = 1 \). The structure zeros which arise from the polynomial part of the 6-j coefficient are indeed polynomial zeros of degree one. All these polynomial zeros of degree one are accounted for by the simple factors:

\[ (1 - \delta_{X,Y} \delta_{n,1}), \]

where

\[ X = (B_u+n)(B_v+n)(\beta_0+1), \]

\[ = (\beta_u - \alpha_0)(\beta_v - \alpha_0)(\beta_0 + 1). \]

and

\[ Y = A_p A_q A_r, \]

\[ = (\beta_0 - \alpha_p)(\beta_0 - \alpha_q)(\beta_0 - \alpha_r). \]

Koozekanani and Biedenharn (1974) calculated the 6-j coefficient for arguments \( j_i, l_i \leq 18.5 \) for \( i = 1, 2, 3 \) and found its polynomial zeros. Using the symmetries of the 6-j coefficient, they ordered the arguments \( j_1, j_2, j_3, l_1, l_2, l_3 \) in a speedometric fashion with \( j_1 \) the slowest varying and \( l_3 \) the most rapidly changing variable. The vanishing values of the 6-j coefficient were calculated by a computer program which resorted to the use of numbers decomposed into powers of prime factors. Bowick (1976) used the Regge symmetries (Appendix 2) of the 6-j coefficient to obtain a reduced listing of the Regge-inequivalent polynomial zeros for arguments \( \leq 18.5 \).

On the other hand, we have used a simple program which checks for the multiplicative factor (20) being equal to zero, to find all the polynomial zeros of degree one, for \( j_i, l_i \leq 18.5 \) for \( i = 1, 2, 3 \). Having separated the majority of the zeros, we sorted out the remaining polynomial zeros tabulated by Koozekanani and Biedenharn (1974), according to their degree given by \( n = \beta_0 - \alpha_0 \). These tables can be found in Srinivasa...
Rao and Rajeswari (1984b). From our classified tables we conclude that 1174 out of 1420 polynomial zeros tabulated by Koozekanani and Biedenharn (1974) are polynomial zeros of degree one, revealed by (20) which is a consequence of the exact binomial expansion (19).

We have also shown (Srinivasa Rao 1985) that of the 12 generic or Regge-inequivalent zeros which have been explained to be due to either triangle rule violation for quasi-spin (Koozekanani and Biedenharn 1974), or vanishing of fractional parentage coefficients in the atomic g-shell (Judd 1970), or realizations of exceptional Lie algebras $G_2$, $F_4$ and $E_6$ (Van der Jeugt et al. 1983; Van den Berghe et al. 1984)—eleven are trivial polynomial zeros of degree one. We also conjectured (Srinivasa Rao 1984) that, in principle, one can find closed form formulas for the polynomial zeros of the 3-j and the 6-j coefficients, provided we look upon these coefficients as generalized hypergeometric functions of unit argument, which are analytic, and extend the method of Siewert and Burniston (1972) to determine zeros of analytic functions to the case of analytic $pFq(l)$s.

Acknowledgement

The author would like to record his indebtedness to Dr R R Ramanna for his constant encouragement and interest and he deems it a pleasure and a privilege to contribute this article to the special issue of this journal to mark his 60th birthday.

Appendix 1

Regge (1958) symmetries for the 3-j coefficient, written down explicitly (Srinivasa Rao 1978) are:

\[
\begin{pmatrix}
  j_1 & j_2 & j_3 \\
  m_1 & m_2 & m_3 \\
\end{pmatrix} = \begin{pmatrix}
  j_1 & \frac{1}{2} (j_2 + j_3 + m_1) & \frac{1}{2} (j_2 + j_3 - m_1) \\
  j_2 - j_3 & \frac{1}{2} (j_3 - j_2 + m_1) + m_2 & \frac{1}{2} (j_3 - j_2 + m_1) + m_3 \\
\end{pmatrix}
\]

\[
= \begin{pmatrix}
  \frac{1}{2} (j_1 + j_3 + m_2) & j_2 & \frac{1}{2} (j_1 + j_3 - m_2) \\
  \frac{1}{2} (j_3 - j_1 + m_2) + m_1 & j_1 - j_3 & \frac{1}{2} (j_3 - j_1 + m_2) + m_3 \\
\end{pmatrix}
\]

\[
= \begin{pmatrix}
  \frac{1}{2} (j_1 + j_2 - m_3) & \frac{1}{2} (j_1 + j_2 + m_3) & j_3 \\
  \frac{1}{2} (j_1 - j_2 + m_3) + m_1 & \frac{1}{2} (j_1 - j_2 + m_3) + m_2 & j_2 - j_1 \\
\end{pmatrix}
\]

\[
= \begin{pmatrix}
  \frac{1}{2} (j_1 + j_2 - m_3) & \frac{1}{2} (j_2 + j_3 - m_1) & \frac{1}{2} (j_1 + j_3 - m_2) \\
  j_3 - \frac{1}{2} (j_1 + j_2 + m_3) & j_1 - \frac{1}{2} (j_2 + j_3 + m_1) & j_2 - \frac{1}{2} (j_1 + j_3 + m_2) \\
\end{pmatrix}
\]

\[
= \begin{pmatrix}
  \frac{1}{2} (j_1 + j_2 + m_3) & \frac{1}{2} (j_2 + j_3 + m_1) & \frac{1}{2} (j_1 + j_3 + m_2) \\
  \frac{1}{2} (j_1 + j_2 - m_3) - j_3 & \frac{1}{2} (j_2 + j_3 - m_1) - j_1 & \frac{1}{2} (j_1 + j_3 - m_2) - j_2 \\
\end{pmatrix}
\]

Only the first and fourth of these symmetries are essentially new, the others can be obtained from these and the column permutations and space reflection symmetries. For an interpretation of the symmetry of the Clebsch-Gordan coefficients discovered by Regge, refer Bincer (1970).
Appendix 2

The Regge (1959) symmetries for the Racah coefficient are:

\[
\begin{align*}
\{ a b e \} &= \{ a \ \frac{1}{2} (b + c + e - f) \ \frac{1}{2} (b - c + e + f) \} \\
\{ d c f \} &= \{ b \ \frac{1}{2} (b + c - e + f) \ \frac{1}{2} (-b + c + e + f) \}
\end{align*}
\]

\[
\frac{1}{2} (a - d + e + f) \ b \ \frac{1}{2} (a + d + e - f) \\
\frac{1}{2} (-a + d + e + f) \ c \ \frac{1}{2} (a + d - e + f)
\]

\[
\frac{1}{2} (a + b + c - d) \ \frac{1}{2} (a + b - c + d) \ e \\
\frac{1}{2} (-a + b + c + d) \ \frac{1}{2} (a - b + c + d) \ f
\]

\[
\frac{1}{2} (b + c + e - f) \ \frac{1}{2} (a - d + e + f) \ \frac{1}{2} (a + b - c + d) \\
\frac{1}{2} (b + c - e + f) \ \frac{1}{2} (-a + d + e + f) \ \frac{1}{2} (a - b + c + d)
\]

\[
\frac{1}{2} (b - c + e + f) \ \frac{1}{2} (a + d + e - f) \ \frac{1}{2} (a + b + c - d) \\
\frac{1}{2} (-b + c + e + f) \ \frac{1}{2} (a + d - e + f) \ \frac{1}{2} (-a + b + c + d)
\]

References

Bowick M J 1976 Regge symmetries and null 3-j and 6-j symbols, Thesis (Christchurch: University of Canterbury)
Judd B R 1970 in Topics in atomic and nuclear theory (Christchurch: New Zealand: Caxton Press)
Koozekanani S H and Biedenharn L C 1974 Rev. Mex. Fis. 23 327
Racah G 1942 Phys. Rev. 62 438
Raynal J 1979 J. Math. Phys. 20 2398
Regge T 1958 Nuovo. Cimento. 10 544
Regge T 1959 Nuovo. Cimento. 11 116
Rose M E 1955 Multipole fields (New York: John Wiley)
Sato M 1955 Prog. Theor. Phys. 13 405
Smorodinskii Ya A and Shelepina L A 1972 Sov. Phys. Usp. 15 1
Srinivasa Rao K and Rajeswari V 1984a *J. Phys.* **A17** L243
Wills J G 1971 *Computer Phys. Commun.* **2** 381
Modified perturbation series for the anharmonic oscillator using linearization technique

NAZAKAT ULLAH
Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

Abstract. The linearization technique of random phase approximation is applied to the anharmonic oscillator to find a modified perturbation series. It is shown that for the anharmonic term $\lambda x^4$, the ground state energy $E_0$ up to the second order of perturbation is given by $E_0 = (35/48) (3/4)^{1/3} \lambda^{1/3}$ as $\lambda \to \infty$.

Keywords. Perturbation series; anharmonic oscillator; random phase approximation.

PACS No. 03 65

1. Introduction

From the very beginning of quantum mechanics it has been known that there are very few eigenvalue problems which can be exactly solved analytically. For physical problems in reality one can either find the eigenfunctions and eigenvalues numerically or resort to some kind of perturbation treatment (Thouless 1972). In the straightforward application of perturbation theory to many problems it is found that successive terms in perturbation expansion do not converge, and hence, various modified forms of expansion have been proposed from time to time. In the present work we shall use the linearization technique of Random Phase Approximation (RPA) (Rowe 1970) to find a modified perturbation series. Extensive literature on the treatment of anharmonic oscillators (Hioe and Montroll 1975; Singh et al 1978) being available, we have chosen the anharmonic oscillator problem for comparison of the modified perturbation series obtained using the linearization technique.

We describe the formulation in § 2 and present the conclusions in § 3.

2. Formulation

Let us consider a hamiltonian $H$ given by

$$H = -\frac{1}{2} \frac{\partial^2}{\partial x^2} + \frac{1}{2} x^2 + \lambda x^4,$$  \hspace{1cm} (1)

where $\lambda x^4, \lambda > 0,$ is the anharmonic perturbation. Applying the ordinary perturbation theory treating the anharmonic term $\lambda x^4$ as perturbation, the series start diverging with the increase in the coupling constant $\lambda$.

The essential idea of the linearization technique is to approximate the term $x^4$ as $\langle x^2 \rangle x^2$, where $\langle x^2 \rangle$ is the expectation value of $x^2$ with respect to some vacuum state.
The perturbation expansion is then applied to the left-over term. We therefore split $H$ in the following way:

$$H = H_0 + H_1,$$  \hspace{1cm} (2)

where

$$H_0 = -\frac{1}{2} \frac{\partial^2}{\partial x^2} + \frac{1}{2} \left(1 + 2\beta\right)x^2,$$  \hspace{1cm} (3)

and

$$H_1 = \lambda x^4 - \beta x^2,$$  \hspace{1cm} (4)

where $\beta$ is an unknown constant to be determined.

The eigenfunctions of $H_0$ can immediately be written as

$$\psi_n = N_n H_n \left(\sqrt{b} x\right) \exp\left(-\frac{1}{2} bx^2\right),$$  \hspace{1cm} (5a)

$$b = \sqrt{1 + 2\beta},$$  \hspace{1cm} (5b)

$$N_n = \left[\frac{\sqrt{b}}{\sqrt{\pi \ 2^n \ n!}}\right]^{1/2}$$  \hspace{1cm} (5c)

The next problem is to determine the unknown constant $\beta$. From our knowledge of manybody physics (Nesbet 1955) we know that the largest configuration interaction effects arise from the lowest configurations. We, therefore, determine $\beta$ by postulating that the matrix element of $H_1$ between $\psi_0$ and $\psi_2$ must vanish. This gives the following cubic equation to determine $\beta$:

$$\beta^2 \left(1 + 2\beta\right) - 9\lambda^2 = 0.$$  \hspace{1cm} (6)

Simple calculations now give the contributions to the ground state energy $E_0$ in various orders of perturbation. We present these results in table 1, where the results obtained using ordinary perturbation theory are also shown.

We calculate the ground state energy for $\lambda = 1$, which gives us an idea about the accuracy of the present formulation. We find from (6) that for this case $\beta = 3/2$. From table 1, we get the value of $E_0$ up to the second order of perturbation to be $E_0 = 0.801$. The value of $E_0$ as given by Hioe and Montroll (1975) using the Bargman representation is $E_0 = 0.804$. Thus our $E_0$ has an error of 0.4%, which is quite small.

The other interesting case is the one in which the coupling constant $\lambda$ becomes very

<table>
<thead>
<tr>
<th>Table 1. Contributions of various orders of perturbation $E_0^{(n)}$ to ground state energy.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contribution to ground state energy</td>
</tr>
<tr>
<td>ordinary$^a$</td>
</tr>
<tr>
<td>modified$^b$</td>
</tr>
</tbody>
</table>

$^a$ ordinary: from ordinary perturbation theory; $^b$ modified: by use of the present formulation
Modified perturbation series

large. For this case (6) gives the following value of $\beta$:

$$\beta = (9/2)^{1/3} \lambda^{2/3}. \tag{7}$$

$\lambda \to \infty$

Using this value of $\beta$ in table 1, we find that the ground state energy $E_0$ up to the second order of perturbation is given by

$$E_0 = \frac{35}{48} \left(\frac{3}{4}\right)^{1/3} \lambda^{1/3},$$

$$E_0 = 0.66 \lambda^{1/3}. \tag{8}$$

This is quite close to the value of $E_0 = 0.67 \lambda^{1/3}$ as given by Hioe and Montroll (1975).

3. Concluding remarks

We have shown how the linearization technique of RPA can be used to obtain a modified perturbation expansion. The results up to the second order in perturbation agree very well with the accurate values given by Hioe and Montroll (1975). The formulation proposed in the manuscript is fairly general and can be applied to other physical problems. Thus, though it may not always be possible to linearize the perturbation operator it may be modified by introducing suitable expectation values to obtain a modified series.

References

Rowe D J 1970 Nuclear collective motion (London: Metheuen) ch. 14
On the Feigenbaum-Cvitanović equation in the theory of chaotic behaviour

VIRENDRA SINGH
Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

Abstract. We propose an analytic perturbative approach for the determination of the Feigenbaum-Cvitanović function and the universal parameter $a$ occurring in the Feigenbaum scenario of period doubling for approach to chaotic behaviour. We apply the method to the case $Z = 2$ where $Z$ is the order of the unique local maximum of the nonlinear map. Our third order approximation gives $a = 2.5000$ as compared to "exact" numerical value $a = 2.5029 \ldots$ We also obtain a reasonably accurate value of the Feigenbaum-Cvitanović function.

Keywords. Feigenbaum-Cvitanović equation; chaotic behaviour; analytic perturbative approach.

PACS No. 05.45; 02.30; 02.90

1. Introduction

The discovery by M Feigenbaum of universal quantitative behaviour, in the maps of an unit interval onto itself, has generated enormous excitement as it opens up the possibility of quantitatively understanding turbulence and other chaotic natural phenomenon. In Feigenbaum scenario a system follows period doubling route to chaos (Feigenbaum 1978, 1979, 1980).

In this theory asymptotically, at each period doubling the separation between any two corresponding adjacent elements of a periodic attractor is scaled by a constant ratio $\alpha$. This leads to the existence of a function $g(x)$ which reproduces itself under the mapping except for the relevant scaling. This function satisfies the equation,

$$-\alpha g(g(x/\alpha)) = g(x),$$

(1)

and is normalised by

$$g(0) = 1.$$  

(2)

We shall refer to this equation and $g(x)$ as Feigenbaum-Cvitanović equation and function respectively (Feigenbaum 1978). For a generalisation to period n-tuplings of this equation we refer to Cvitanović-Myrheim (1983). Equation (1) together with normalisation (2) determines both the functional form of $g(x)$ and $\alpha$. We shall be interested in the solution which has a local maximum at $x = 0$ of the order $Z$. Such a solution is unique (Feigenbaum 1980, p. 15; Collet et al 1980).

The universal function $g(x)$ and $\alpha$ do, of course, depend on the order of local maximum i.e. if the transformation $T$ considered is

$$T: x \rightarrow x' = \lambda(1 - a|x|^2),$$

where $\lambda$ and $a$ are constants.
then they depend on $Z$. For $Z = 2$, which is the most interesting case, we have

$$\alpha = 2.502907875 \ldots$$

The function $g(x)$ has also been determined numerically for this case (Feigenbaum 1979).

Another important universal constant $\delta$, which determines the rate of period-doublings, is given by

$$\delta = 4.6692016 \ldots$$

for $Z = 2$. A consideration of asymptotic period-doubling leads to the equation

$$-\alpha[h(g(x/\alpha)) + g'(g(x/\alpha))h(x/\alpha)] = \delta h(x)$$

$$\delta > 1$$

for a determination of the universal function $h(x)$ and universal constant $\delta$. Again numerical determination of $h(x)$ and $\delta$ has been carried out (Feigenbaum 1979).

In view of the importance of these functions it would be desirable to have an analytical approach for their determination as opposed to one involving pure numerical computation. We propose an analytical perturbative scheme towards this purpose in the present paper. It is a systematic procedure and can be improved successively. Among other attempts to determine $\alpha$, but not the $g(x)$, by analytic approximations, we may mention those of Helleman (1983) and Hu and Mao (1982) among others.

### 2. Basic equations and method

Replacing $x$ by $\alpha x$ we get

$$g(\alpha x) + \alpha g[g(x)] = 0,$$

$$g(0) = 1.$$  \hfill (4)  \hfill (5)

Let

$$g(x) = 1 + p(x),$$

$$p(x) = \sum_{n=1}^{\infty} p_n x^n,$$

$$y = |x|^Z \quad p_1 \neq 0,$$

where $Z$ is the order of local maximum of the transformation.

Further let

$$p(\alpha x) = \sum_{n=1}^{\infty} C_n [p(x)]^n.$$  \hfill (6)  \hfill (7)  \hfill (9)

Using (6) and (9) in (4) we obtain

$$1 + \sum_{n=1}^{\infty} C_n [p(x)]^n + \alpha g[1 + p(x)] = 0.$$  \hfill (10)

Equating the coefficients of $[p(x)]^n$ to zero in (10) we obtain

$$1 + \alpha g(1) = 0.$$  \hfill (11)
and
\[ n! C_n + \alpha f^{(n)}(1) = 0 \quad (n = 1, 2, \ldots). \] (12)

On the other hand, if we expand both sides of (9) in powers of \( y \) and equate the coefficients we obtain \((n = 1, 2, \ldots)\)
\[ p_n \beta^n = \sum_{l, m_1, m_2, \ldots, m_n \geq 1} C_l p_{m_1} p_{m_2} \cdots p_{m_n} \delta_{m_1 + m_2 + \ldots + m_n = n} \] (13)
where \( \beta = |\alpha|^2 \).

We display first few of these equations.
\[ p_1 \beta = C_1 p_1, \] (14)
\[ p_2 \beta^2 = C_1 p_2 + C_2 p_1^2, \] (15)
\[ p_3 \beta^3 = C_1 p_3 + C_2 (2p_1 p_2) + C_3 p_1^3, \] (16)
\[ p_4 \beta^4 = C_1 p_4 + C_2 (2p_1 p_3 + p_2^2) + C_3 (3p_1^2 p_2) + C_4 p_1^4, \] (17)

We note that since \( p_1 \neq 0, \)
\[ C_1 = |\alpha|^2. \] (18)

It is also useful to note that
\[ f^{(i)}(1) = \sum_{r=1}^{\infty} \frac{(rZ)!}{r!(rZ-l)!} P_r \] (19)
for \( l = 1, 2, 3, \ldots, \) and
\[ f(1) = 1 + \sum_{r=1}^{\infty} P_r. \] (20)

If we combine (11), (13), (19) and (20), and further define
\[ p_n \alpha^n = S_n |\alpha|^Z \] (21)
\[ n = 1, 2, \ldots. \]

We obtain the basic equations to be solved
\[ \frac{1}{\alpha} + 1 + |\alpha|^Z \sum_{r=1}^{\infty} \frac{S_r}{\alpha^r} = 0 \] (22)
and
\[ S_n + \sum_{U} \frac{(rZ)!}{r! (rZ-l)!} \cdot \frac{S_r}{\alpha^{r-1}} \cdot \frac{S_{m_1} S_{m_2} \cdots S_{m_i}}{\alpha^{(n-l)}} = 0 \]
for \( n \geq 1. \) (23)

The summation in (23) is over the set \( U \) given by
\[ r \geq 1, \quad l \geq 1, \]
\[ m_1 \geq 1, \quad m_2 \geq 1, \quad \ldots, \quad m_i \geq 1 \]
and
\[ m_1 + m_2 + \ldots + m_i = n. \]
We now note that (23) for $n = 1$ leads to

$$S_1 + \sum_{r \geq 1}^{\infty} \left( \frac{rZ}{1} \right) S_r S_1 = 0$$

and since $S_1 \neq 0$ we get

$$\frac{1}{Z} + \sum_{r \geq 1}^{\infty} \frac{r S_r}{Z^{r-1}} = 0. \quad (24)$$

In fact using this equation we can simplify the remaining equations (23) for $n \geq 2$ to read

$$S_n \left( 1 - \frac{1}{|z|^Z} \right) + \sum_{l \geq 2}^{Z(n-1)} \sum_{r \geq 1}^{\infty} \left( \frac{rZ}{l} \right) S_r \times \frac{\sum S_{m_1} S_{m_2} \ldots S_{m_n} \delta_{m_1 + m_2 + \ldots + m_n}}{|z|^Z(n-1)} = 0$$

for $n = 2, 3, \ldots \quad (25)$

Equations (22) and (24) and (25) together constitute a system of coupled nonlinear equations to determine $\alpha$ and $S_1, S_2, \ldots$. We regard (22) as the eigenvalue equation for $\alpha$ while (24) and (25) are to be used to determine $S_1, S_2, \ldots$ in terms of $\alpha$.

These equations are, however, highly nonlinear and coupled. One needs a systematic procedure for their solution. We now notice that (24)-(25) allow us to conclude that, as $\alpha \to \infty$, we have a solution for which $S_n$ tends to constants $S_{n, 0}$, i.e.

$$S_n(\alpha) \xrightarrow{\alpha \to \infty} S_{n, 0}.$$ 

In particular $S_n$'s admit, as $\alpha \to \infty$, an expansion in the inverse powers of $\alpha$, of the form, for $Z$ a nonrational number,

$$S_n(\alpha) = \sum_{p, q = 0, 1, 2, \ldots} S_{n, p, q} \left( \frac{1}{\alpha} \right)^{n, p, q},$$

where $S_{n, p, q}$ are numerical constants. When $Z$ is integral or a rational number simpler expansions can be written down.

We propose to use such $\alpha \to \infty$ expansions and use them to $S_n(\alpha)$ in terms of $\alpha$ by using (24) and (25). These expansions for $S_n(\alpha)$ when substituted in (22) would give us the eigenvalue equation for the determination of $\alpha$.

It is known that as $Z \to 1$, $\alpha$ does tend to infinity. We thus expect this procedure to clearly work when $Z \to 1$ (Collet et al 1980; Derrida et al 1978, 1979). We will see in the next section, where we apply it to the physically most interesting case $i.e. Z = 2$ that the method is still quite effective.

3. Application of the method for $Z = 2$

We shall now apply our method for the $Z = 2$ case. In this case we can use the simpler expansion given by
Feigenbaum-Cvitanović equation

\[ S_n(\alpha) = \sum_{m=0, 1, 2, \ldots} \frac{S_{n,m}}{\alpha^m}. \]  

(26)

Using (24) we obtain, as \( \alpha \to \infty \)

\[ S_1(\alpha) \to -1/2. \]

(27)

Similarly from (5) and using (27) we obtain the following leading behaviours, as \( \alpha \to \infty \)

\[ S_2 \to \frac{1}{8}, \quad S_3 \to \frac{1}{16\alpha}, \quad S_4 \to -\frac{1}{128\alpha} \text{ etc.} \]

(28)

Proceeding further and equating powers of \( \alpha \) on both sides of (24) and (25) we obtain

\[
O = S_{1,1} + 2S_{2,0} \\
O = S_{2,1} + 3S_{1,0}S_{1,1} + 6(S_{1,0})^2S_{2,0}, \quad O = S_{1,2} + 2S_{2,1} + 3S_{3,0} \\
O = S_{2,2} - S_{2,0} + (S_{1,0})^2[S_{1,2} + 6S_{2,1} + 15S_{3,0}] \\
+ 2S_{1,0}S_{1,1}[S_{1,1} + 6S_{2,0}] + [2S_{1,0}S_{1,2} + S_{1,1}]S_{1,0}, \\
O = S_{1,3} + 2S_{2,2} + 3S_{3,1} + 4S_{4,0} \text{ etc.} \]

(29)

Combining the information contained in (28) with that contained in (29) we obtain

\[
S_{1,0} = -1/2; \\
S_{1,1} = -1/4, \quad S_{2,0} = 1/8; \\
S_{1,2} = 0, \quad S_{2,1} = 0, \quad S_{3,0} = 0; \\
S_{1,3} = -1/4, \quad S_{2,2} = 1/32, \quad S_{3,1} = 1/16, \quad S_{4,0} = 0; \text{ etc.} \]

(30)

Using these coefficients we obtain the approximation

\[
g(\alpha) = 1 - \left( \frac{\alpha}{2} + \frac{1}{4} + \frac{1}{4\alpha^2} + \ldots \right)x^2 + \left( \frac{1}{8} + \frac{1}{32\alpha^2} + \ldots \right)x^4 \\
+ \left( \frac{1}{16\alpha^2} + \ldots \right)x^6 + \ldots
\]

(31)

where all the coefficients of the powers of \( x^2 \) have been kept to the order \( 1/\alpha^2 \).

The eigenvalue equation is

\[ J(\alpha) = 0, \]

(32)

where \( J(\alpha) \equiv - \frac{1}{\alpha} + 1 + \alpha \sum_{n=1}^{\infty} \frac{S_n}{\alpha^{n-1}} \) and we have

\[ J(\alpha) = - \frac{\alpha}{2} + \frac{7}{8} + \frac{1}{\alpha} - \frac{5}{32\alpha^2} + O(1/\alpha^3). \]

(33)

To this order in approximation we obtain from (32) using (33)

\[ \alpha = 5/2 = 2.5000 \]

(34)

to be compared with the result of the “exact” numerical computation \( \alpha = 2.5029 \ldots \).
Using the approximate value \( a = 5/2 \) we get from the expression (31),
\[
g(x) \approx 1 - 1.5400x^2 + 0.1300x^4 + 0.0100x^6 + \ldots
\]  
We also quote, for comparison, the result of the "exact" numerical computation (Feigenbaum 1979)
\[
g(x) \approx 1 - 1.5276x^2 + 0.1048x^4 + 0.0267x^6 - 0.0035x^8 + \ldots
\] where we have kept only the four figures after the decimal sign.

### 4. An exact solution

An exact solution of (4) is given by
\[
g(x) = (1 + |x|^Z)^{1/Z},
\]
\[
a = \frac{1}{2^{1/Z}}.
\]
This solution is, however, of no interest for the Feigenbaum scenario for the onset of the chaotic behaviour since \( g(x) \) has a local minimum rather than local maximum at \( x = 0 \). It however implies that the eigenvalue equation for \( a \) should have a root given by (38). For \( Z = 2 \) we should thus have a root at \( a = -1/\sqrt{2} \approx -0.707 \ldots \) for the exact eigenvalue equation. Our approximate eigenvalue equation given by (32) and (33) i.e.
\[
O = \frac{a + 7}{2} + \frac{1}{\alpha} - \frac{5}{32x^2}
\]
has a negative root at \( a = -1/\sqrt{17} + 3/8 \approx -0.8904 \) which presumably a reflection of the root \( a \approx -0.707 \ldots \) for the exact equation.

### 5. Concluding remarks

We thus see that the present method produces reasonably accurate results. Since the method is a systematic one the accuracy can be improved by taking higher order approximations.

The real utility of the method would be to further allow us to calculate the \( Z \)-dependance of the universal parameter \( a \). Moreover whenever one needs \( g(x) \) as an input, such as for example in the equations for \( h(x) \), it is useful to have such analytic approximations to \( g(x) \) as we have obtained here. These and similar investigations will be reported later.

### References

Feigenbaum–Cvitanović equation

Feigenbaum M J 1980 *Los Alamos Science* **1** 4
Quantum chaos and fluctuation properties of regular and irregular spectra

JITENDRA C PARIKH* and VIJAY B SHEOREY
Physical Research Laboratory, Navrangpura, Ahmedabad 380009, India
* Present address: Department of Physics, Texas A&M University, College Station, Texas 77843, USA (on leave of absence from P.R.L., Ahmedabad)

Abstract. Fluctuation properties of the regular and irregular energy levels for the Hénon-Heiles Hamiltonian are examined. The spacing distributions and the calculated values of the $A_3$-statistic show that there is no difference in the short range correlation properties of these spectra. Remarkably, the $A_3$ values agree with the results of random matrix theory.

Keywords. Quantum chaos; fluctuation properties; regular spectra; irregular spectra; spectral signature; spacing distribution.

PACS No. 05-40; 24-60; 46-60

1. Introduction

During the last few decades much progress has been made in understanding the (time) evolution of classical dynamical systems in phase space. In particular the studies have provided a classification scheme for dynamical systems based on how "regular" or "erratic" the motion is in phase space. As a result with increasing "irregularity" of motion, the systems are in turn labelled periodic, quasi-periodic, ergodic, mixing and chaotic ($K$-systems). An account of the earlier work on dynamical systems has been given by Khinchin (1949) and the more recent formal developments are described by Arnold and Avez (1968), Ornstein (1974), Arnold (1978) and Krylov (1978). Ford (1973) reviewed these developments in a more physical and intuitive way.

In this paper our interest is in "regular" and chaotic motion. It is therefore worth recalling that, for regular (periodic and quasi-periodic) motion the system remains confined to a restricted region (in general invariant tori) of the total available phase space. Also recall that for a chaotic system the motion is over the entire available phase space; it is extremely "erratic" and it shows random behaviour. This occurs in spite of the fact that the system is fully (all degrees of freedom) described and the equations of motion governing its evolution are completely deterministic. Although many aspects of "regular" and chaotic motion in classical systems are fairly well understood, the way in which they are expressed in the corresponding quantum systems are not always equally clear. (For reviews on "quantum chaos" see Zaslavsky 1981 and Bohigas and Giannoni 1984).

The purpose of this paper is to discuss one aspect of the "quantum chaos" problem that has been extensively studied in recent years. More precisely, we describe how "regular" and chaotic motion in a bound classical Hamiltonian system will be reflected in the discrete energy spectrum of the corresponding quantum system. The suggestions
in this respect have included (i) examining properties of individual energy levels (Percival 1973) and, (ii) examining statistical (fluctuation) properties (Zaslavsky et al 1974; Zaslavsky 1977) of a whole set of levels.

In view of these suggestions, we summarize in § 2 results of earlier studies on spectral signatures which characterize quantum “regular” levels and quantum “irregular” (chaotic) levels. We also point out (what in our view is) an omission in all studies involving fluctuation properties. It turns out that this has a bearing on signatures which are a consequence of statistical properties of energy levels.

Section 3 contains a brief description of our calculations and results for the Hénon-Heiles (Hénon and Heiles 1964) model Hamiltonian. A summary and concluding remarks are presented in § 4.

2. Spectral signatures of regular and irregular levels

Following Percival's work (Percival 1973) there are several studies proposing signatures in the energy spectrum that will distinguish between the “regular” levels (classically quasi-periodic motion in phase space) and the “irregular” levels (classically chaotic motion). The basic conclusion is that “irregular” levels correspond to large second differences (Pomphrey 1974) or “avoided” crossings (Noid et al 1980) in energy. This is not the case for “regular” levels. Further, it appears (Pullen and Edmonds 1981) that large second differences correspond to “avoided” crossings.

Another suggestion (Zaslavsky et al 1974; Zaslavsky 1977) has been to study the distribution of energy spacings between neighbouring levels. It has been shown that, for the “regular” spectrum, this distribution (Berry and Tabor 1977) is of the Poisson form, and for the “irregular” spectrum (McDonald et al 1979; Casati et al 1980; Pechukas 1983; Bohigas et al 1984; Haller et al 1984; Seligman et al 1984) it is essentially of the Wigner form due to repulsion between neighbouring levels belonging to the same (exact) symmetry.

It should be stressed however that for any interacting system neighbouring levels belonging to the same exact symmetry will repel each other. Since for “irregular” levels there are no other quantum numbers (constants of motion) except energy, level repulsion is to be expected in their spacing distribution. For “regular” motion if the system (having \( f \) degrees of freedom) is integrable there are \((f - 1)\) other constants of motion besides energy. Even if the system is not integrable, for regular motion it follows from Kolmogorov-Arnold-Moser (KAM) theorem (see e.g. Arnold and Avez 1968) that the phase space trajectory will be on a deformed torus. This suggests that in this case also there ought to be in addition to energy \((f - 1)\) other constants of motion. It is not clear at all how to determine these additional constants of motion for a non-integrable system. All the same it is crucial that they (actually the corresponding quantum numbers) be considered when obtaining the fluctuation properties of “regular” levels. If this is not done, the resulting properties would very likely be a consequence of a superposition of uncorrelated level sequences. We believe that the Poisson spacing distribution for the “regular” levels is an artefact arising from just such a superposition of uncorrelated energy levels. This is because no other quantum numbers except energy are taken account of in the earlier studies.
3. Fluctuation properties of Hénon-Heiles Hamiltonian

Keeping in view these considerations, we determine in this section the fluctuation properties of the energy eigenvalues of the modified Hénon-Heiles (Hénon and Heiles 1964; Pullen and Edmonds 1981) Hamiltonian. The reason for choosing this Hamiltonian is that it has only two degrees of freedom, but more important has been the fact that in our knowledge it is the Hamiltonian which has been most extensively studied in connection with classical and quantum chaotic behaviour.

For this Hamiltonian, we have obtained the nearest neighbour spacing distribution and the $\Delta_3$-statistic of Dyson and Mehta (1963). For evaluating these properties we have very closely followed the approach of Haq et al (1982). We would like to determine (i) if there is a difference between the fluctuation properties of the “regular” and the “irregular” spectrum, and (ii) how these properties compare with the prediction of random matrix theory.

As we shall see, we are able to provide unambiguous answers using both the spacing distribution and the $\Delta_3$-statistic.

We briefly describe our calculations next. The model Hamiltonian (Pullen and Edmonds 1981) is

$$H = \frac{1}{2} (p_x^2 + p_y^2 + x^2 + y^2) + \alpha(x^2y - 1/3 y^3),$$

where $\alpha$ is the strength parameter. Note that the anharmonic term has a 3-fold symmetry ($C_{3v}$ point group) (Pullen and Edmonds 1981). We have separately calculated the eigenvalues for each of the 3-irreducible representations (labelled $A_1$, $A_2$, $E$) for $\alpha = 0.088$. This particular value was used by Pomphrey (1974) and Pullen and Edmonds (1981). The diagonalization procedure used (reducing to tridiagonal form followed by bisection) was the same that Pullen and Edmonds followed. The number of eigenvalues we obtain for each symmetry below the escape energy ($1/6\alpha^2 = 21.52$) is shown in table 1. The convergence of the eigenvalues was good to four significant places when enlarging the basis space. For each symmetry we identified the “regular” and the “irregular” levels by considering table 2 of Pullen and Edmonds. It turns out then that most of the levels below $E = E_c$ ($E_c = 14.5-17.5$) belong to the “regular” spectrum and those with $E > E_c$ belong to the “irregular” spectrum. $E_c$ is different for the three symmetries. We then find that out a total of 171 levels, 87 belong to the regular part. The remaining 84 levels characterize irregular levels.

The fluctuations (Porter 1965; Mehta 1967; Brody et al 1981) in levels are (by definition) departures from a smooth uniform spectrum. Since for our model $H$ the

<table>
<thead>
<tr>
<th>Symmetry type</th>
<th>Number of eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>50</td>
</tr>
<tr>
<td>$A_2$</td>
<td>37</td>
</tr>
<tr>
<td>$E$</td>
<td>84</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>171</strong></td>
</tr>
</tbody>
</table>
level density is not a constant, it is essential to "unfold" (Brody et al. 1981) the spectrum and map it to one with constant density. The mapping we have chosen is

\[ F(E) = a + bE + cE^2. \]  

(2)

This is a natural choice because the quadratic term is the distribution function (continuous approximation) for the unperturbed oscillator. The parameters \(a, b, c\) were determined separately for each symmetry by a best fit to the calculated energy levels.

We very briefly discuss next the basic definition and physical meaning of the spacing distribution \(p(x)\) and the \(\Delta_3\)-statistic. \(p(x)\) \(dx\) is the probability for the spacing between two levels, with no levels lying between them, to take a value between \(x\) and \(dx\). Note that \(x\) is measured in units of constant mean spacing and \(\int_0^\infty p(x)dx = 1\). As described in detail elsewhere (Bohigas and Giannoni 1975) the expressions for \(\Delta_3\) can be reduced (for purposes of calculations) to

\[ \Delta_3(n) = \frac{N^2}{16} \left[ \frac{1}{n^2} \left( \sum_{i=1}^N E_i \right)^2 + \frac{3N}{2n^2} \left( \sum_{i=1}^N E_i^2 \right) - \frac{3}{n^4} \left( \sum_{i=1}^N E_i^2 \right)^2 \right] + \frac{1}{n} \sum_{i=1}^N (N - 2i + 1) E_i. \]  

(3)

In (3) \(n\) denotes a fixed energy interval in units of (constant) mean spacing and \(N\) is the actual number of energy levels in this interval. The \(N\) levels have energies \(E_i\) \((i = 1, \ldots, N)\). Physically, \(\Delta_3(n)\) provides a measure of departure of the exact eigenvalue distribution function (staircase curve) from a straight line. It should be recalled that (Dyson and Mehta 1963) for the gaussian orthogonal ensemble \((\text{goe})\) of real symmetric matrices, the mean \(\Delta_3(n)\) over the ensemble is known for \(n \to \infty\). For finite values of \(n\) it has been obtained by Bohigas et al (1983) using Monte-Carlo calculations.

The results of our calculations for \(p(x)\) and \(\Delta_3\) are discussed next.

### 3.1 Spacing distributions

When \(\alpha = 0\), the system (equation (1)) is integrable and one has only a "regular" spectrum. The energy levels (two-dimensional harmonic oscillator) are labelled by the total number of quanta \(n\) and "angular momentum" \(l\), where \(n = 0, 1, 2, \ldots\), and for a fixed \(n\), \(l = -n, -n + 2, \ldots, n - 2, n\). Note that all levels for a fixed \(n\) are degenerate. Clearly, the spacing distribution of levels with a fixed \(l\) is a \(\delta\)-function, which shows extreme level repulsion and rigidity. One would obtain a Poisson distribution if one superimposes many uncorrelated \(l\)-sequences.

When \(\alpha \neq 0\), the system is not integrable and classically one has either regular motion \((\text{KAM})\) regime or chaotic motion in phase space. In the former case, the motion in phase space is confined to a two-dimensional surface (deformed torus). Hence, there ought to be another "local" constant of motion (quantum number) besides energy for the "regular" levels. This quantum number should be included in the analysis of fluctuation properties of the "regular" levels. If this is done, then since the system \((\text{C}_{3v} \text{ oscillator basis})\) is an interacting system, one would expect level repulsion for both "regular" and "irregular" levels. Our results (without the additional quantum number for "regular" levels) are shown in figure 1. They show level repulsion for both "regular" and "irregular" levels. These results are different from those of Haller et al (1984) and Seligman et al (1984) for the "regular" levels. One possibility for the difference is that
Figure 1. Spacing distributions for regular levels (figure 1a) and irregular levels (figure 1b). Histograms denote the calculated distributions for the Hénon-Heiles potential. Smooth curves denote the theoretical (Wigner) distribution \( p(x) = \frac{\pi}{2} x \exp\left(-\frac{\pi}{4} x^2\right) \).

compared to our case, they may have a much larger number of uncorrelated sequences in their calculations.

3.2 \( \Delta_3 \)-statistic

We have evaluated \( \Delta_3(\bar{n}) \) for \( 0 < \bar{n} \leq 10 \) by separately taking the spectral average over the "regular" and the "irregular" levels belonging to each symmetry, and then taking the mean ("ensemble" average) over the three symmetry subspaces. The results are shown in table 2.

The result of \( \Delta_3(\bar{n}) \) for \( \bar{n} \leq 5 \) clearly shows no discernible difference between the values for the "regular" and "irregular" spectrum. For values of \( \bar{n} > 5 \) the difference is \( \approx 10\% \) but then the statistics are not as good as those for \( \bar{n} < 5 \). The agreement with the theoretical prediction of GOE is also very good for small \( \bar{n} \) and deteriorates to 15–20\% difference for larger \( \bar{n} \). There is however no agreement between the calculated \( \Delta_3(\bar{n}) \) and the Poisson distribution values.

In view of the numerical results, for \( \Delta_3 \), it seems reasonable to conclude that whenever we have been able to carry out sharpest (best statistics) comparisons, there is
Table 2. Values of $\Delta_3(\bar{n})$ for regular, irregular, GOE (from Bohigas et al 1983) and Poisson spectra ($\Delta_3(\bar{n}) = \bar{n}/15$).

<table>
<thead>
<tr>
<th>$\bar{n}$</th>
<th>regular</th>
<th>irregular</th>
<th>GOE</th>
<th>Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.0169</td>
<td>0.0169</td>
<td>0.0165</td>
<td>0.0167</td>
</tr>
<tr>
<td>0.50</td>
<td>0.0326</td>
<td>0.0334</td>
<td>0.0325</td>
<td>0.0333</td>
</tr>
<tr>
<td>0.75</td>
<td>0.0475</td>
<td>0.0489</td>
<td>0.0471</td>
<td>0.0500</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0611</td>
<td>0.0630</td>
<td>0.0605</td>
<td>0.0667</td>
</tr>
<tr>
<td>2.0</td>
<td>0.1127</td>
<td>0.1147</td>
<td>0.1023</td>
<td>0.1333</td>
</tr>
<tr>
<td>3.0</td>
<td>0.1535</td>
<td>0.1506</td>
<td>0.1320</td>
<td>0.2000</td>
</tr>
<tr>
<td>4.0</td>
<td>0.1807</td>
<td>0.1770</td>
<td>0.1549</td>
<td>0.2667</td>
</tr>
<tr>
<td>5.0</td>
<td>0.1927</td>
<td>0.1953</td>
<td>0.1735</td>
<td>0.3333</td>
</tr>
<tr>
<td>6.0</td>
<td>0.1977</td>
<td>0.2100</td>
<td>0.1893</td>
<td>0.4000</td>
</tr>
<tr>
<td>7.0</td>
<td>0.2025</td>
<td>0.2204</td>
<td>0.2028</td>
<td>0.4667</td>
</tr>
<tr>
<td>8.0</td>
<td>0.2077</td>
<td>0.2267</td>
<td>0.2148</td>
<td>0.5333</td>
</tr>
<tr>
<td>9.0</td>
<td>0.2120</td>
<td>0.2340</td>
<td>0.2255</td>
<td>0.6000</td>
</tr>
<tr>
<td>10.0</td>
<td>0.2140</td>
<td>0.2379</td>
<td>0.2356</td>
<td>0.6667</td>
</tr>
</tbody>
</table>

no difference between the fluctuation properties of regular, irregular, and GOE spectra. Thus such fluctuation measures (statistic) are not likely to provide a signature for distinguishing between the “regular” and the “irregular” spectra. It is also remarkable that a system with just two degrees of freedom leads to fluctuation properties that are identical to GOE—essentially a parameter-free theory originally proposed for a system with a larger number of degrees of freedom.

4. Summary and concluding remarks

Our results show no significant difference in the fluctuation properties of regular, irregular and GOE spectra. As stated before, it is not clear why our results for “regular” spectra do not agree with those of Haller et al (1984) and Seligman et al (1984). While our statistics are small compared to theirs, it seems unlikely that our conclusions would be altered in a significant manner with improved statistics. For this purpose further calculations are currently in progress. The agreement between the results for “irregular” spectra and GOE is consistent with earlier results and seems to suggest (Bohigas et al 1984) universality in the behaviour of fluctuation properties—independent of the number of degrees of freedom.

Acknowledgements

We wish to thank O Bohigas, R Haq, M J Giannoni and A Pandey for many discussions. We would also like to thank O Bohigas, R Haq and A Pandey for providing us with some of their computer programs and results of Monte-Carlo calculations prior to publication.
References

Arnold V I and Avez A 1968 *Ergodic problems of classical mechanics* (New York: Benjamin)
Bohigas O and Giannoni M J 1984 Lectures delivered at Workshop on *Mathematical and computational methods in nuclear physics* October 1983 Granada, Spain—to be published in *Lecture Notes in Physics* (Springer-Verlag)
Haller et al 1984 *Phys. Rev. Lett.* 52 1665
Hénon M and Heiles C 1964 *Astron. J.* 69 73
Khinchin A I 1949 *Mathematical foundations of statistical mechanics* (New York: Dover)
McDonald S W et al 1979 *Phys. Rev. Lett.* 42 1189
Pechukas P 1983 *Phys Rev. Lett.* 51 943
Percival I C 1973 *J. Phys.* B6 L229
Pomphrey N 1974 *J. Phys.* B7 1909
Pullen R A and Edmonds A R 1981 *J. Phys.* A14 L319
Zaslavsky G M 1977 *Sov. Phys. JETP* 46 1094
Self-energy effect in a relativistic bound state problem

S N BISWAS and ASHOK GOYAL
Department of Physics, University of Delhi, Delhi 110007, India

Abstract. A modification of the Wick-Cutkosky equation for the relativistic bound state of two scalar particles interacting through the exchange of a massless scalar field within the ladder approximation has been considered by incorporating the self-energy diagrams in the integral kernel. An exact analytical solution of the equation is obtained at vanishing total energy and it is shown that the self-energy effects generally diminish the eigenvalues in agreement with the findings of Li et al., who, however solved the equation numerically for the case of massive scalar exchange. An additional feature of the modified equation is that it preserves the 0(5) symmetry at zero total energy as was first noted by Cutkosky for the scalar bound state equation without self-energy effects.

Keywords. Bethe-Salpeter equation; relativistic bound state problem; self-energy effects.

PACS No. 11.10

1. Introduction

The covariant wave equation originally formulated by Bethe and Salpeter (1951) more than thirty years ago still remains the main tool to study the bound state problem in the context of quantum field theory. In momentum space it is given by an integral equation with integral kernel determined by four-point irreducible Feynman diagrams as well as by self-energy diagrams. In the absence of a complete knowledge of all these, the ladder approximation is generally used. In this context the studies of Wick (1954) and Cutkosky (1954) for a scalar Bethe-Salpeter equation (bs equation) have revealed many interesting consequences of a relativistic bound state problem. Later Blanckenbecler and Sugar (1966) have pointed out the effect of unitarity condition on the bs equation and in particular Levine and Wright (1967) have shown that the unitarity relation namely $\sigma_{\text{total}} \geq \sigma_{\text{inelastic}}$ can be violated in the inhomogeneous bs equation describing scattering processes if only the ladder diagrams are taken into the integral kernel. Inclusion of self-energy diagrams, do, however, remove this difficulty. Recently Li et al (1980) have considered the influence of the self-energy diagrams on the solutions of the bound state equation involving scalar particles. Briefly, they considered a scalar bs equation with the interaction chosen as $\mathcal{H}(x) = g\phi_1^*(x)\phi_1(x)\phi_2(x)$, where $\phi_1$ is a complex scalar field with mass $m$ and $\phi_2$ is a neutral scalar field with mass $\mu$. In the second-order the integral kernel of the equation has the form

$$\frac{i\lambda}{\pi^2 (p-p')^2 + \mu^2} - \lambda h(p)(p_1^2 + m^2)(p_2^2 + m^2)\delta^4(p-p'),$$

where $h(p)$ stands for the self-energy contribution to the kernel and $p_1$ and $p_2$ are respectively $p + \frac{1}{2}P$ and $-p + \frac{1}{2}P$. $P_\mu$ is the centre-of-mass four-momentum of the composite system. It is well known that for nonvanishing $\mu$ the scalar bs equation, even
with the first term of the kernel given by (1), cannot be solved analytically. Li et al (1980) adopted a comprehensive numerical study of the problem with the integral kernel given in (1). Their calculations show that self-energy diagrams in general diminish the eigenvalues and also alter the wave functions slightly.

In this note we point out that with a little modification of the kernel, exact analytic solution of the problem exists for some given values of total energy and that the properties of the eigenvalues with respect to the coupling parameter can be easily determined. Li et al (1980) have discussed in detail through their numerical analysis the eigenvalues of the coupling parameter for various non-vanishing values of \( \mu \) and for a given set of values of the energy parameter \( \eta = E/2m \) where \( E \) is the fourth component of \( P_\mu \). In order to obtain analytic solutions we modify the kernel as

\[
\frac{i \lambda}{\pi^2} \frac{1}{(p-p')^2} - \lambda h(p)(p_1^2 + m^2)(p_2^2 + m^2) \delta^4(p - p').
\]

We thus consider the situation when only massless scalar particles are exchanged, however, the mass of the neutral scalar contributing to the self-energy expression \( h(p) \) need not be zero and can be taken at will. The equation we thus consider is a simple generalization of the original Wick (1954)-Cutkosky (1954) equation together with an additional term in the kernel due to self-energy contribution to second order in coupling strength \( g(\lambda \approx g^2) \).

2. Separability of the modified Wick’s equation

The modified Wick’s equation which we like to solve reads

\[
(p_1^2 + m^2)(p_2^2 + m^2) \chi(p) = -\frac{i \lambda}{\pi^2} \int \frac{\chi(p') d^4 p'}{(p-p')^2} + \lambda h(p)(p_1^2 + m^2)(p_2^2 + m^2) \chi(p),
\]

where both momentum \( p \) and \( p' \) are in Euclidean space, and the self-energy function \( h(p) \) is given by

\[
h(p) = \int_{(m+\mu)^2}^{\infty} \frac{(p_1^2 + m^2)[\sigma^2 - (m + \mu)^2]^{1/2}[\sigma^2 - (m - \mu)^2]^{1/2}}{\sigma^2(\sigma^2 - m^2)(p_1^2 + \sigma^2)} d\sigma^2 + (p_1 \to p_2)
\]

We rewrite (3) as

\[
\psi(p) = -\frac{i \lambda}{\pi^2} \int \frac{d^4 p' \psi(p')}{(p-p')^2(p_1^2 + m^2)(p_2^2 + m^2)[1 - \lambda h(p')]} \]

where \( p_1' \) and \( p_2' \) are respectively equal to \( p' + (E/2) \) and \( -p' + (E/2) \) and \( \psi(p) = [1 - \lambda h(p)](p_1^2 + m^2)(p_2^2 + m^2) \chi(p) \). Hereafter, we will assume that the 3-vector part of \( P_\mu \) is a null-vector for simplicity. Equation (5) can also be written as a differential equation i.e.

\[
\Box \psi(p) = -4 \lambda \frac{\psi(p)}{(p_1^2 + m^2)(p_2^2 + m^2)} \cdot \frac{1}{1 - \lambda h(p)},
\]

and

\[
\Box = (\partial^2/\partial p_1^2) + (\partial^2/\partial p_2^2).
\]
We now show that (6) is separable in terms of new (Green 1957; Biswas 1967) variables defined by

\[ p_1 = p_s \sin \theta \cos \phi, \]
\[ p_2 = p_s \sin \theta \sin \phi, \]
\[ p_3 = p_s \cos \theta, \]

and further, \( p_s = C \sin \eta / (\cosh \alpha - \cos \eta) \) and \( p_4 = C \sinh \alpha / (\cosh \alpha - \cos \eta) \).

The whole energy-momentum plane is contained in the region \( 0 \leq \theta \leq \pi, 0 \leq \phi \leq 2\pi \) and \(-\infty \leq \alpha \leq \infty \) and the parameter \( C \) is given by

\[ C^2 = (E/2)^2 + m^2 \]

In terms of these new variables, the self-energy function \( \psi(p) \) becomes a function of the variable \( \alpha \) only. For instance, introducing a quantity \( c' \) with

\[ C'^2 = (E/2)^2 + \sigma^2. \]

The function \( \psi(p) \) can be written as,

\[ \psi(p) = f(\alpha) g(\eta) Y_{lm}(\theta, \phi)/p_3. \]

Substituting (13) in (6) we obtain after separation of variables the following two equation for \( g(\eta) \) and \( f(\alpha) \) namely,

\[ \frac{d^2 g}{d\eta^2} + \left\{ n^2 - \frac{l(l+1)}{\sin^2 \eta} \right\} g = 0, \]

and

\[ \frac{d^2 f}{d\alpha^2} - \left\{ n^2 - \frac{\lambda}{C^2 \cosh^2 \alpha - \frac{E^2}{4} \sinh^2 \alpha} \cdot \frac{1}{1 - \lambda \Delta(m, \mu, \tanh \alpha)} \right\} f = 0, \]

where \( n \) is a separation constant.

We thus see that the modified Wick-Cutkosky equation which includes the effect of self-energy insertions in the integral kernel of the \( bs \) equation still admits separable solution. In §3 we discuss the eigen solution of the problem and concentrate on the particular case \( E = 0 \) which affords a simple exact solution of (15) determining the eigenvalues.

3. Solution at \( E = 0 \)

At \( E = 0 \) we note from (11) that \( \Delta(m, \mu, \tanh \alpha) \) assumes a particularly simple form: Here \( \Delta(m, \mu, \tanh \alpha) \) becomes independent of \( \alpha \) and \( \Delta(m, \mu, \tanh \alpha) \) reduces to a constant
depending on m and \( \mu \) only which we denote by \( \Delta \), i.e.

\[
\Delta(m, \mu, \tanh \alpha)_{E=0} = \Delta,
\]

and (15) then reduces to

\[
\frac{d^2 f(\alpha)}{d\alpha^2} - \left\{ n^2 - \frac{\lambda/(1 - \lambda\Delta)}{m^2 \cosh^2 \alpha} \right\} f(\alpha) = 0.
\]

The boundary conditions on \( g(\eta) \) and \( f(\alpha) \) in the new coordinates are obtained by substituting them in the original integral (9) which is satisfied if \( g(0) = g(\pi) \) and \( f(\alpha) \) vanishes at large \( \alpha \) and is an even function of \( \alpha \). Further \( f \) and \( g \) must be finite within the given ranges as mentioned at the end of (8).

The acceptable solutions of (14) consistent with the above mentioned boundary conditions are the well-known Gegenbauer polynomials,

\[
g(\eta) = \text{constant} \times \sin^{l+1}(\eta) \cdot C_{\pi - \eta}^{l+1}(\cos \eta).
\]

Equation (17) determines the energy eigenvalues of the bound states. Here, for a given \( E \), the equation, however, determines the spectrum of values of \( \lambda \). If we transform (17) in terms of a new variable \( \chi \) where \( \cos \chi = \tanh \alpha \), the solution of (17) can be easily obtained as

\[
f = \text{constant} \times \sin^n(\chi) C_{\pi - n}^{n+1/2}(\cos \chi),
\]

where the integer \( N \) is related to the eigenvalue \( \lambda \) by

\[
\frac{\lambda}{m^2(1 - \Delta \lambda)} = N(N + 1).
\]

This result is to be compared with the eigen spectrum of \( \lambda \) when the effect of self-energy is not present, which is obtained from (20) by setting \( \Delta = 0 \) on the left-hand side. The relevant relation is

\[
\lambda/m^2 = N(N + 1).
\]

Thus when \( E = 0 \) and the scalar particles interact through ladder diagrams via the exchange of massless scalar particle the self-energy diagrams for \( \Delta > 0 \) will diminish the eigenvalue. This result is in agreement with the finding of Li et al (1980), who, however, solved the problem numerically for a massive scalar field exchange. The eigenfunctions in the present case when we consider the massless scalar particle exchange remain identical both for equations with or without the inclusion of self-energy effects. Li et al (1980) found a slight change in the nature of the eigenfunctions in these two cases.

### 4. Other remarks

If we collect the solutions at \( E = 0 \) of \( g(\eta) \) and \( f(\alpha) \) from (18) and (19) we see from (13) that the Bethe-Salpeter wave-function \( \psi \) takes the form

\[
\psi \approx Y_{Nnlm}
\]

where \( Y_{Nnlm} \) are the spherical harmonics in the 5-dimension euclidean space and are given by
Self-energy effect

\[ Y_{n\ell m} = (\sin \chi)^n C_{\frac{n+1}{2}}(\cos \eta)^{\ell+1} C_{\frac{n-1}{2}} \cdot Y_{\ell m}(\theta, \phi), \]

where \( N > n > \ell + 1, \ell > |m| \) and the degeneracy of \( Y_{n\ell m} \) is \( \frac{1}{2} N(N+1)(2N+1) \). In other words the equation even in the presence of self-energy effects preserves the 0(5) symmetry (Biswas 1967) originally discovered by Cutkosky (1954) when the self-energy effect was not included.

At \( E \neq 0 \) the energy-eigenvalue equation (17) can be recast in the following form

\[
(1 - Z^2) \frac{d^2 G}{dZ^2} + 2(n - 1)Z \frac{dG}{dZ} + \left\{ \frac{\lambda/m^2}{(1 - E^2 + E^2Z^2)} \cdot \frac{1}{(1 - \lambda \Delta(m, \mu, Z^2))} \right\} G = n(n - 1)G.
\]

Equation (20) reduces to Cutkosky's (1954) equation when \( \Delta \) is set to zero. The equation cannot be solved exactly except at special limits. Both the eigenfunctions and eigenvalues will be largely altered due to the presence of the factor \([1 - \lambda \Delta(Z^2)]^{-1}\) in the third term which makes the equation more singular than the standard Huen's equation (Erdelyi 1942, 1944).

References

Bethe H A and Salpeter E E 1951 Phys. Rev. 84 1232
Biswas S N 1967 J. Math. Phys. 8 1109
Blanckenbecler R and Sugar R 1966 Phys. Rev. 142 1051
Cutkosky R E 1954 Phys. Rev. 96 1135
Erdelyi A 1942 Quart. J. Math. 13 107
Erdelyi A 1944 Quart. J. Math 15 64
Green H S 1957 Nuovo Cimento 5 866
Wick G C 1954 Phys. Rev. 96 1124
Angular momentum and isospin properties of chiral fields

KLAUS GOEKE and J N URBANO*
Institut fur Kernphysik, Kernforschungsanlage Juelich, D-5170 Juelich, West Germany
Institut fur Theor. Kernphysik, Universitat Bonn, D-5300 Bonn, West Germany
* Departmento de Fisica da Universidade, P-3000 Coimbra, Portugal

Abstract. In a recently suggested variational quantum field theoretical approach the angular momentum and isospin properties of the pion field surrounding a quark bag are investigated using the Lagrangian of the Cloudy Bag Model.

Keywords. Angular momentum; isospin properties; chiral fields; variational quantum field; mean field approximation.

PACS No. 1235

1. Introduction

In the last few years various methods have become popular in order to solve chiral bag, chiral solition and Skyrme models for the nucleon and delta system. Most of these procedures (Chodos and Thorn 1975; Brown and Rho 1979; Vento et al 1980; Birse and Banerjee 1984; Kalbermann and Eisenberg 1984; Jackson and Rho 1983; Adkins et al 1983) seek solutions assuming from the start the so-called hedgehog form (Skyrme 1962) because the formal simplification allows an exact nonperturbative treatment of these models in the mean field approximation. Actually in a recent paper some justification of the hedgehog assumption has been given (Urbano and Goeke 1984c). There it has been shown that the hedgehog is one of a set of shapes of the pion mean field, coupled to the corresponding quark structure, which minimizes the energy of the Cloudy Bag Hamiltonian in the nucleon delta sector. The proof has been obtained in a recently formulated variational quantum field theoretical approach (vQF), whose details can be found in Urbano and Goeke (1984a, b). Unfortunately, the hedgehog approximation for the bag-pion system violates rotational and isospin symmetries. This means that the corresponding state is not an eigenstate of the operators of total angular momentum $J$ and total isospin $T$. Thus it is important to investigate some of the $J$- and $T$-properties of the hedgehog and related solutions of the chiral bag system. This is the objective of this paper and the formalism used is the vQF-theory.

The paper is constructed as follows. Section 2 reviews the vQF-theory with particular emphasis on its mean field aspects, §3 investigates some angular momentum and isospin properties of the hedgehog solution. The corresponding projection formalism is developed in §4 with some numerical results for simplified quark-pion structures. A summary and an outlook are given in §5.

2. The vQF-theory

The detailed formulation of the vQF-approach can be found in Urbano and Goeke (1984a, b). In the present section we review briefly the formalism as far as it is needed for
the following. All considerations are presently done using the Lagrangian of the Cloudy Bag Model (Théberge et al 1980). This one is written in the form

$$\hat{H} = \int d^3r \left( -i \sum_{a=1}^{3} \partial_{a} q_{a}^{\dagger}(r) \alpha \cdot \nabla q_{a}(r) + B \right) \theta_{e}$$

$$+ \sum_{j=1}^{3} \int d^3k \alpha_{j}(k) a_{j}(k)$$

$$+ \frac{i}{2f} \int d^3r \sum_{a=1}^{3} q_{a}(r) \gamma_{5} \tau^{(a)} \cdot \hat{\phi}(r) q_{a}(r) \Delta_{a}$$

(1)

Here $\hat{\phi}(r)$ is the quantized pion field operator

$$\hat{\phi}_{j}(r) = \frac{1}{(2\pi)^{3/2}} \int \frac{d^3k}{\sqrt{2\omega(k)}} \left[ a_{j}(k) \exp(ik \cdot r) + a_{j}^{\dagger}(k) \exp(-ik \cdot r) \right]$$

(2)

and the $a_{j}^{\dagger}(k) [a_{j}(k)]$ create (annihilate) a pion with momentum $k$ and cartesian isospin component $j$. The $q_{a}(r)$ is the quark field operator, where the index $a$ indicates the colour.

In the nucleon-delta sector the quarks occupy the s-state of the bag with the familiar wave function

$$\psi_{a}(r) = \frac{N}{4\pi} \left( \frac{j_{0}(\Omega r/R)}{ij_{1}(\Omega r/R)\sigma^{(a)} \cdot r} \right) |\chi_{a} \rangle,$$

(3)

where $|\chi_{a} \rangle$ is the spin-flavour state of the quark with colour $a$. Apparently the most general form of $|\chi_{a} \rangle$ in the $u$-$d$ sector is given by

$$|\chi_{a}(\alpha) \rangle = N(\alpha) \{ |u, \uparrow \rangle + \alpha_{2} |u, \downarrow \rangle + \alpha_{3} |d, \uparrow \rangle + \alpha_{4} |d, \down \rangle \},$$

(4)

where $u, d$ indicate up- and down-quarks and the arrows the orientation of the spin component. The $N(\alpha)$ is the normalization constant and the $\alpha$'s are to be treated as variational parameters for the quark structure of the baryon. Due to symmetry the $\alpha$'s are the same for all three colours.

The total baryon is assumed to consist of the quark core given by (3), and the pion cloud described by a coherent state

$$|\xi \rangle = N(\xi) \exp \left[ \sum_{j} \int d^3k \xi_{j}(k) a_{j}^{\dagger}(k) \right] |0 \rangle.$$

(5)

Here the $N(\xi)$ is a normalization factor and the field amplitudes, $\xi_{j}(k)$, are to be treated as variational parameters for the pion cloud.

Using (3) and (4) the total trial wave function of the baryon reads

$$|\psi(\alpha, \xi) \rangle = \left( \prod_{a} \psi_{a}(r) \right) |\xi \rangle.$$

(6)

For a fully quantum mechanical treatment the $|\psi(\alpha, \xi) \rangle$ has to be projected on good spin and isospin quantum numbers:

$$|\psi_{JT}(\alpha, \xi) \rangle = \frac{P_{j} P_{T} |\psi(\alpha, \xi) \rangle}{\langle \psi(\alpha, \xi) | P_{j} P_{T} |\psi(\alpha, \xi) \rangle^{1/2}}.$$

(7)
Here $P_j$ and $P_T$ are projection operators introduced into nuclear physics by Peierls and Yoccoz. For a general intrinsic solution (6) with $\sigma \cdot \mathbf{k}$ the spin-flavour structure (4) these projections are complicated since no simplifying symmetry relations can be used. Thus projection techniques are presently only used for special spin-flavour configurations as reported in Urbano and Goeke (1984b) and in §5. Here we first investigate the properties of the intrinsic solution corresponding to the mean field approximation.

3. Mean field approximation

In the mean field approximation the energy is given by the expectation value of $H$ between the states (6), yielding

$$
E(\alpha, \xi) = E_{\text{MIT}} + \sum_{j=1}^{3} \int d^3k \omega(k) \xi_j^*(k) \xi_j(k)
+ g \sum_{j=1}^{3} \int \frac{d^3k}{\sqrt{2\omega(k)}} \left[ i\rho(k) \xi_j(k) \sum_{a=1}^{3} \langle \chi_a | \sigma^{(a)} \cdot \mathbf{k} \tau_j^{(a)} | \chi_a \rangle \right] + \text{c.c.},
$$

(8)

where $E_{\text{MIT}} = 3\Omega/R + 4\pi B/3R^3 - Z/R$ is the MIT-model energy, $\rho(k) = 3j_1(kR)/kR$ is the Fourier transform of the pion source density, and $g = \Omega/(6f(\Omega - 1)(2\pi)^{3/2})$ is a renormalized coupling constant.

The variation of (8) with regard to $\xi_j^*(k)$ yields $\delta E / \delta \xi_j^*(k) = 0$ and hence

$$
\xi_j(k) = \frac{ig\rho(k)}{\omega(k)\sqrt{2\omega(k)}} \sum_{a=1}^{3} \langle \chi_a | \sigma^{(a)} \cdot \mathbf{k} \tau_j^{(a)} | \chi_a \rangle.
$$

(9)

Inserting this into (8) the total energy becomes

$$
E(\alpha) = E_{\text{MIT}} - 6\pi g^2 A(\alpha) \int_0^{\infty} \frac{k^4 \omega^2(k)}{\omega^2(k)} \, dk
$$

(10)

with

$$
A(\alpha) = \sum_{i,j=1}^{3} | \langle \chi(\alpha) | \sigma_i \tau_j | \chi(\alpha) \rangle |^2.
$$

(11)

The $E(\alpha)$ has a minimum whenever $A(\alpha)$ reaches a maximum. As shown in Urbano and Goeke (1984c), evaluating $A(\alpha)$ using (4) yields

$$
A(\alpha) = 3 - 2N^4(\alpha)( | \alpha_1 |^2 + | \alpha_2 |^2 - | \alpha_3 |^2 - | \alpha_4 |^2 )^2 + 4 | \alpha_1 \alpha_3^* + \alpha_2 \alpha_4^* |^2.
$$

(12)

One sees immediately that $A(\alpha) \leq 3$. The maximum value of $A(\alpha)$ is reached taking, for instance, $\alpha_1 = \alpha_4 = 0$ and $\alpha_2 = - \alpha_3 = 1$. This particular choice corresponds to the quark spin-isospin state

$$
| \chi_h \rangle = \frac{1}{\sqrt{2}} \{ | u, \downarrow \rangle - | d, \uparrow \rangle \},
$$

(13)

Inserting (13) into (9) yields
\[ \xi_h(k) = -i \frac{6g \rho(k)}{(2\omega(k))^{3/2}} (k_x \delta_{j1} + k_y \delta_{j2} + k_z \delta_{j3}). \] (14)

For the corresponding pion mean field,
\[ \phi_j(r) = \langle \xi_h | \hat{\phi}_j(r) | \xi_h \rangle, \] (15)

one obtains after an explicit calculation
\[ \phi_1(r) = \frac{x}{r} G(r), \quad \phi_2(r) = \frac{y}{r} G(r), \quad \phi_3(r) = \frac{z}{r} G(r). \] (16)

with
\[ G(r) = 3g(2/\pi)^{1/2} \int dk k^3 \rho(k) \left[ \omega(k) \right]^{3/2} j_1(kr). \] (17)

Both (13) and (16) show the well-known and always assumed hedgehog properties. Apparently they arise from the variational principle formulated above, \( \text{ie} \ \text{vqf}. \)

4. Angular momentum and isospin properties of the hedgehog solution

The hedgehog solution \( |\psi_h \rangle \) with \( \chi \) of (13) and \( \xi \) of (14) satisfies for \( i = x, y, z \) the equations
\[ (J_i^{(q)} + T_i^{(q)}) |\psi_h \rangle = 0 \] (18)

and
\[ (J_i^{(n)} + T_i^{(n)}) |\psi_h \rangle = 0. \] (19)

The first one (18) is rather trivial to show, if (13) is considered. The second one (19) is more difficult and will be shown explicitly in this section.

The expansion of the coherent state \( |\xi_h \rangle \) in terms of zero boson, one-boson, two-boson states, etc., yields
\[ |\xi_h \rangle = 0 + \sum_j \int d^3k \xi_j(k)a_j^+(k)|0\rangle + \frac{1}{2!} \sum_{jj'} \int d^3k \int d^3k' \xi_j(k)\xi_{j'}(k')a_j^+(k)a_{j'}^+(k')|0\rangle + \ldots. \]

Due to the structure of the coherent state \( |\xi_h \rangle \) it is sufficient to show that
\[ (J_i^{(n)} + T_i^{(n)}) \sum_j \int d^3k \xi_j(k)a_j^+(k)|0\rangle = 0, \] (20)

In a cartesian three-dimensional representation one has for the isospin operators the matrices
\[ T_1^{(n)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad T_2^{(n)} = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \quad T_3^{(n)} = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \]

If one writes \( a_j^+(k)|0\rangle = |k, j\rangle \) and
\[ |k, 1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} |k\rangle, \quad |k, 2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} |k\rangle, \quad |k, 3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} |k\rangle, \]
Angular momentum and isospin properties of chiral fields

one obtains immediately

\[ T_1^{(n)} |k, 1 \rangle = 0 \quad T_1^{(n)} |k, 2 \rangle = i |k, 3 \rangle \quad T_1^{(n)} |k, 3 \rangle = -i |k, 2 \rangle \]
\[ T_2^{(n)} |k, 1 \rangle = -i |k, 3 \rangle \quad T_2^{(n)} |k, 2 \rangle = 0 \quad T_2^{(n)} |k, 3 \rangle = i |k, 1 \rangle \]
\[ T_3^{(n)} |k, 1 \rangle = i |k, 2 \rangle \quad T_3^{(n)} |k, 2 \rangle = -i |k, 1 \rangle \quad T_3^{(n)} |k, 3 \rangle = 0. \]

Thus one gets eq. (14)

\[ T_1^n \int d^3k \xi_1(k) a_1^+(k) |0\rangle = \int \frac{6g\rho(k)}{[2\omega(k)]^{3/2}} \{ -k_y |k, 3 \rangle + k_z |k, 2 \rangle \} d^3k. \] (21)

The operator \( J_k^n \) can generally be written as

\[ J_k^n = \sum_{m,n} \epsilon_{kmn} \hat{q}_m \hat{p}_n, \] (22)

where \( \hat{q} \) and \( \hat{p} \) are coordinate and momentum operators. This yields in the momentum representation for example

\[ J^n_1 = i \int d^3k |k\rangle \frac{\partial}{\partial k_y} k_z \langle k| - i \int d^3k |k\rangle \frac{\partial}{\partial k_z} k_y \langle k|. \] (23)

Hence one obtains

\[ J^n_1 \int d^3k' \xi_1(k') a_1^+(k') |0\rangle = i 6g \int d^3k' \frac{\rho(k')}{[2\omega(k')]^{3/2}} \times \int d^3k |k\rangle \frac{\partial}{\partial k_y} k_z \{ -k_y' \langle k|k', 1 \rangle - k_z' \langle k|k', 2 \rangle - k_z' \langle k|k', 3 \rangle \} \]
\[ - \int d^3k |k\rangle \frac{\partial}{\partial k_z} k_y \{ -k_y' \langle k|k', 1 \rangle - k_y' \langle k|k', 2 \rangle - k_z' \langle k|k', 3 \rangle \}, \]

giving

\[ J^n_1 \int d^3k' \xi_1(k') a_1^+(k') |0\rangle = -6g \int d^3k \frac{\rho(k)}{[2\omega(k)]^{3/2}} \times \left\{ |k, 1 \rangle \frac{\partial}{\partial k_y} k_z (-k_x) + |k, 2 \rangle \frac{\partial}{\partial k_y} k_z (-k_y) + |k, 3 \rangle \frac{\partial}{\partial k_x} k_z (-k_z) \right\} \]
\[ - |k, 1 \rangle \frac{\partial}{\partial k_z} k_z (-k_x) - |k, 2 \rangle \frac{\partial}{\partial k_z} k_y (-k_y) - |k, 3 \rangle \frac{\partial}{\partial k_z} k_y (-k_z) \right\}. \]

The explicit evaluation yields now

\[ J^n_1 \int d^3k' \xi_1(k') a_1^+(k') |0\rangle = 6g \int d^3k \frac{\rho(k)}{[2\omega(k)]^{3/2}} \left[ -k_z |k, 2 \rangle + k_y |k, 3 \rangle \right]. \] (24)

Comparison of (24) with (21) shows that we have proven the assertion for the considered special case \( i = 1 \) in (10). The proof for the other cases is similar.

From (8) and (9) there follows the important fact that a decomposition of the total hedgehog state \( |\psi_n\rangle \) in terms of states with good spin \( J \) and isospin \( T \) yields

\[ |\psi_n\rangle = \sum_j a_j |J, T = J\rangle. \] (25)
Thus, in order to include quantum correlations, only a projection on one of the quantum numbers, involving only one set of Euler angles, is sufficient for a hedgehog state. However, this simplification does not hold if one first projects on states $J, T$ and then performs a variation. In such a case there is no reason why $|\psi\rangle$ should not have components with $J \neq T$.

5. Projection techniques

A proper way to proceed would be now to perform, first, a projection on good $J$- and $T$-quantum numbers according to (7) and second, to vary the corresponding energy into its minimum. This is a rather complicated procedure which has not been investigated yet. A simpler way would consist in performing the projections after the variation, i.e., to perform a projection of a hedgehog state. Even this is complicated since a hedgehog state has nonaxial components in spin and isospin space. In order to explore the projection techniques a simpler case is considered here, where the intrinsic three-quark states are coupled (Urbano and Goeke 1984b; Theberge et al 1980) to the ground state of a proton with spin up, $|N_{1/2}^+\rangle$ mixed with a state $|\Delta_{1/2}^+\rangle$. Thus the bare nucleon state is assumed to be

$$|BN\rangle = \cos \alpha |N_{1/2}^+\rangle + \sin \alpha |\Delta_{1/2}^+\rangle. \tag{26}$$

For such a state one obtains the energy

$$E_{JT}(\alpha, R) = \int_0^\pi d\beta \sin \beta \int_0^\pi d\tilde{\beta} \sin \tilde{\beta} d_{1/2}^T 1/2(\beta)T_{1/2}^T 1/2(\tilde{\beta})h(\beta, \tilde{\beta}) + \int_0^\pi d\beta \sin \beta \int_0^\pi d\tilde{\beta} \sin \tilde{\beta} d_{1/2}^T 1/2(\beta)T_{1/2}^T 1/2(\tilde{\beta})n(\beta, \tilde{\beta}), \tag{27}$$

with the overlap kernels

$$h(\alpha; \beta, \tilde{\beta}) = \langle BN | \langle \xi | \exp \{i\beta J_y + i\tilde{\beta} J_3\} | BN \rangle | \xi \rangle,$$

$$n(\alpha; \beta, \tilde{\beta}) = \langle BN | \langle \xi | \exp \{i\beta J_y + i\tilde{\beta} J_3\} | BN \rangle | \xi \rangle. \tag{28}$$

For the evaluation of (28) it is necessary to know the rotated pion field state

$$\exp \{i\beta J_y + i\tilde{\beta} J_3\} | \xi \rangle = \frac{1}{N} \exp \left\{ \sum_j \int d^3 k \xi_j(k; \beta \tilde{\beta}) a_j^+(k) \right\} | 0 \rangle,$$

with

$$\xi_j(k, \beta \tilde{\beta}) = \sum_j \left[ R_3(\tilde{\beta}) \right]_{ij} \xi_j(R^{-1}_y(\beta) k),$$

where $N$ is a normalization constant. Following the formalism of Urbano and Goeke (1984a) some results of the calculations are shown in figure 1. There the mr-energy, the intrinsic energy and the projected energy are contrasted. One realizes that the effect of the projection is biggest at small bag radii where the pion field is strongest. Actually for very small $R$-values one would obtain a collapse of the system if this is not prevented by an additional mechanism. In the present investigation this is done by assuming the $q\bar{q}$-component of the pion to have a finite extension (Urbano and Goeke 1984b) characterized by its radius $\eta_{\pi}$. For a physical pion the $\eta_{\pi}$ is related to the pion decay
Angular momentum and isospin properties of chiral fields

Figure 1. The total nucleon energy is given for various bag radii $R$. Considered are the MIT-model, the VQF-formalism in the mean field approximation (intrinsic) and with angular momentum and isospin projections included (projected). For the last case $\eta_a = 0.17$ fm and $\eta_n = 0$ fm are considered to demonstrate the instability for small radii for vanishing $\eta_n$. The parameters of the calculation are $B^{1/4} = 0.125$ GeV, $g = 14$ GeV$^{-1}$, $Z = 1.75$.

Figure 2. The average number of pions in the intrinsic nucleon state is given in dependence on the mixing angle $\alpha$ for various bag radii $R$. The parameters of the calculation are those of figure 1.

constant and comes out to be $\eta_a = 0.17$ fm. This value is used in the calculations.

An interesting quantity is the average number of pions in the cloud. This is given by

$$N_\pi = \langle \bar{\xi} \left| \sum_{j=1}^{3} \int d^3k a_j^+ (k) a_j (k) \right| \xi \rangle,$$
and for the projected states analogously. Some results are found in figures 2 and 3. Figure 2 shows for various $R$-values the intrinsic number of pions in dependence on the mixing angle $\alpha$. One sees a clear variation showing maximal values at $\alpha$ around 40°. For bag radii of about 1-2 fm only small values of $N_\pi$ are obtained which are at most $N_\pi \approx 0\cdot 2$. The projected results are different. They are displayed in figure 3 for various $\eta_\pi$-values in dependence of $R$. A comparison between both figures shows that the projected values are noticeably larger than the nonprojected ones. This holds even for

![Figure 3](image)

Figure 3. The average number of pions in the projected nucleon state is given in dependence on the bag radius $R$ for various mixing angles $\alpha$. The parameters of the calculation are those of figure 1.

![Figure 4](image)

Figure 4. The self-energy the bag acquires due to the coupling with the pion field. Considered is the intrinsic solution with the parameters of figure 1 for various $\alpha$- and $R$-values.
Angular momentum and isospin properties of chiral fields

Figure 5. The self-energy of the projected state for various $\alpha$- and $R$-values. The parameters are taken from figure 1.

\[ \alpha = 0^\circ \text{ and } R \approx 1.2 \text{ fm} \text{ where the effect of the projections on the energy is rather small.} \]

The self-energy of the system shows a behaviour as it is displayed in figures 4 and 5. It is noticeably larger in magnitude for the projected case. For all the above data it should be noted that due to the finite pion size the effective coupling constant goes to zero for $R \to 0$.

6. Summary

The objective of the present investigation was to illuminate angular momentum and isospin properties of chiral bag models. To this end the variational quantum field theoretical approach (vqf) of Urbano and Goeke was considered. It was first demonstrated that the well known hedgehog structure appears by variational techniques if a Cloudy Bag Hamiltonian is used. It was then shown that this hedgehog allows for simplifications of the projection techniques to be used. Those were then studied in a simplified model showing that particularly the average number of pions in the field is a rather sensitive number. Stability of the system can be guaranteed if a finite size of the pions in the cloud is assumed.

References

Klaus Goeke and J N Urbano

Skryrne T H R 1962 Nucl. Phys. 31 556
Urbano J N and Goeke K 1984b preprint, Jülich
Relativistic harmonic confinement of quarks and gluons

S B KHADKIKAR
Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India

Abstract. A current confinement model for glue-balls is proposed to maintain complete similarity with relativistic harmonic oscillator model (rhm) with Lorentz scalar and vector potentials for quarks. The spurious motion of the centre of confinement is accounted for in these models in exactly the same manner. We shall review some of the successes of the rhm in hadronic masses, magnetic moments of baryons, nucleonic sizes and polarisability and proton-antiproton annihilation, as well as the recent results on glue-balls.

Keywords. Relativistic harmonic confinement; quarks; gluons; glue-balls; hadrons; masses; magnetic moments.

PACS No.

1. Introduction

It is now accepted that strongly interacting particles i.e. hadrons consist of quarks and gluons and the dynamics of their 'colour' is quantum-chromo-dynamics (QCD). The QCD is asymptotically free and one expects it to show infrared slavery or confinement in accord with the empirical fact that only colour singlet objects are seen free. Starting from a confinement model one can derive properties of hadrons and this was the philosophy of early bag models (Bogolinkov 1967, 1972; Chodos et al 1974). These models though relativistically suffered from sharp surface and spurious motions of the centre of confinement (cm). The non-relativistic constituent quark models (Isgur and Carl 1977, 1978, 1979a,b) are not consistent though they could explain many aspects of hadron-spectroscopy. The relativistic harmonic oscillator (Lorentz scalar plus vector) model (rhm) (Khadkikar and Gupta 1983) combined the desirable features of both and was eminently successful in explaining very diverse aspects of hadron spectroscopy and magnetic moments, nucleon polarisability, nucleon anti-nucleon annihilation, etc. We shall review some of the results in the following. QCD predicts that there should be new kinds of particles—colour singlet states of two or more gluons—glue-balls and hybrids of quarks and gluons—meiktons. Here we shall propose a particular scheme for confinement of gluons on similar lines to rhm for quarks. We shall then compare the results to experimental candidates for glue-ball states. In the end we shall present an outlook for future work.

2. rhm for hadrons with 'interaction' effects

The rhm for light hadrons with a constant effective one gluon exchange hyperfine interaction (Khadkikar and Gupta 1983) gave quite an impressive account of hadronic
masses and baryonic magnetic moments. However there was the problem of the pion being too heavy. Also the magnetic moments though better than other model calculations were not in precise agreement with experimental values. This meant that the additional interaction effects on the magnetic moments which were neglected are not insignificant. By using group theoretic methods one can analyse the contribution to the magnetic moments in terms of unit tensors of \((\text{SU}_3 \times \text{SU}_3)\) flavour group. Indeed a single tensor was found so that the discrepancy of the RHM moments from experimental values could be removed. The nature of the tensor is such as to satisfy Marshak-Okubo-Sudarshan sum rule and the trace of the magnetic moment, over the flavour octet is nearly zero. The results obtained (Gupta and Khadkikar 1984) are given in table 1.

Table 1. Magnetic moments of octet baryons with q-moments.

<table>
<thead>
<tr>
<th>Baryons</th>
<th>RHM + T</th>
<th>Experimental</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p)</td>
<td>2.759</td>
<td>2.7928456</td>
<td>2.793</td>
</tr>
<tr>
<td>(n)</td>
<td>-1.894</td>
<td>-1.91304184</td>
<td>-1.913</td>
</tr>
<tr>
<td>(\Lambda)</td>
<td>-0.572</td>
<td>-0.613 \pm 0.004</td>
<td>-0.614</td>
</tr>
<tr>
<td>(\Sigma^+)</td>
<td>2.335</td>
<td>2.38 \pm 0.02</td>
<td>2.388</td>
</tr>
<tr>
<td>(\Xi^0)</td>
<td>0.605</td>
<td>—</td>
<td>0.633</td>
</tr>
<tr>
<td>(\Sigma^-)</td>
<td>-1.126</td>
<td>-1.111 \pm 0.033</td>
<td>-1.123</td>
</tr>
<tr>
<td>(\Xi^*)</td>
<td>-1.177</td>
<td>-1.250 \pm 0.014</td>
<td>-1.245</td>
</tr>
<tr>
<td>(\Xi^-)</td>
<td>-0.637</td>
<td>-0.69 \pm 0.04</td>
<td>-0.685</td>
</tr>
<tr>
<td>(\Lambda \to \Sigma^0)</td>
<td>-1.499</td>
<td>-1.82 \pm 0.26</td>
<td>-1.520</td>
</tr>
</tbody>
</table>

The interaction effects are simulated by a tensor in isospin \(T = d I (2I - 1) + 4I_3 (I - 1)\), \(d\) is the only parameter. We also show a fit when the three quark magnetic moments are treated as free parameters.

Table 2. Calculated masses of light hadrons with renormalised strength in \((\pi, K)\) systems (Khadkikar and Gupta 1984) compared with observations.

<table>
<thead>
<tr>
<th>Hadron</th>
<th>Observed (MeV)</th>
<th>(M_{\text{cal}}) (MeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p/n)</td>
<td>939</td>
<td>939</td>
</tr>
<tr>
<td>(\Lambda)</td>
<td>1116</td>
<td>1115</td>
</tr>
<tr>
<td>(\Sigma)</td>
<td>1189</td>
<td>1181</td>
</tr>
<tr>
<td>(\Xi)</td>
<td>1321</td>
<td>1330</td>
</tr>
<tr>
<td>(\Delta)</td>
<td>1232</td>
<td>1232</td>
</tr>
<tr>
<td>(\Sigma^*)</td>
<td>1385</td>
<td>1375</td>
</tr>
<tr>
<td>(\Xi^*)</td>
<td>1533</td>
<td>1523</td>
</tr>
<tr>
<td>(\Omega^-)</td>
<td>1672</td>
<td>1678</td>
</tr>
<tr>
<td>(\rho)</td>
<td>770</td>
<td>745</td>
</tr>
<tr>
<td>(\kappa^*)</td>
<td>892</td>
<td>901</td>
</tr>
<tr>
<td>(\omega)</td>
<td>783</td>
<td>745</td>
</tr>
<tr>
<td>(\phi)</td>
<td>1019</td>
<td>1068</td>
</tr>
<tr>
<td>(\pi)</td>
<td>139</td>
<td>135</td>
</tr>
<tr>
<td>(K)</td>
<td>495</td>
<td>497</td>
</tr>
</tbody>
</table>
The other problem of the pion being rather heavy arises because the one gluon- 
hyperfine interaction is treated perturbatively and because the interaction is attractive 
in \( \pi \) and \( \rho \) one can show that the diagonalisation of the interaction matrix in \( RHM \) states 
leads to the lowering of the \( \pi \) and \( \rho \) ground states so as to agree with observed masses 
(Gupta and Khadkikar 1984, table 2).

Considering the successes of \( RHM \) which we shall further discuss later, it is tempting to 
construct a gluon confinement model on parallel lines. One should keep in mind that 
the success of \( RHM \) was closely linked with the accounting for spurious CM.

3. A confinement model for glue-balls

QCD predicts the existence of gluonic bound colour singlet states called glue-balls. For, 
short of any exact solutions of QCD as in the case of hadrons, one can proceed to make 
models with confinement as the prime criterion. Dielectric confinement models have 
been in vogue e.g. see Lee 1981). In the harmonic confinement scheme, however, the 
dielectric confinement alone seems to generate tremendous complications especially for 
the colour electric modes (Khadkikar and Vinodkumar 1984) for want of a suitable 
gauge choice. The simplicity is recovered though if one assumes that the confinement is 
due to current (\( J \)) which is proportional to the ‘gluon’ vector field (\( A \)) for each colour,

\[
J = -a^2 r^2 A,
\]

where \( r \) is position vector from the centre of confinement. It is now straightforward to 
write down Maxwell’s equations with current source (1) and solve them in a Lorentz 
gauge to obtain for \( A \)

\[
(-\nabla^2 + a^2 r^2) A = \omega^2 A.
\]

We have to still eliminate the additional (longitudinal) degree of freedom in \( A \). The 
natural choice here is what we might call an oscillator gauge. One can write down (2) in 
oscillator quanta

\[
(a \cdot a^+ + a^+ \cdot a) A = \omega^2 A,
\]

where

\[
a = \frac{1}{\sqrt{2a}} (\nabla + ar),
\]

and \( a^+ \) is the hermitean conjugate. The usual radiation gauge conditions \( \nabla \cdot A = 0 \) is 
now replaced by

\[
a \cdot A = 0
\]

in the oscillator gauge, implying

\[
\nabla \cdot A + ar \cdot A = 0,
\]

and

\[
-ar \cdot A + \frac{\partial \phi}{\partial t} = 0,
\]

where \( \phi \) is the coulomb potential.

It is now seen that in this gauge the scalar modes are nullified by the radial 
component of the vector gluons. The electric (\( E \)) and magnetic (\( M \)) modes are now 
given by the conditions:
\[ a^+ \cdot A = 0 \ (M), \] (5)

and

\[ L \cdot A = (a^+ \times a) \cdot A = 0 \ (E). \] (5a)

The \( M \) and \( E \) modes then are given in terms of a scalar oscillator function \( \phi_{nlm} \) satisfying the eigenvalue equation the same as \( A \) as follows:

\[ A_M = L \phi_{nlm}, \]

and

\[ A_E = (a - a^+ a \cdot a) \phi_{(n+1)lm}. \] (6)

Thus \( n = 1 \) and \( n = 0 \) are the lowest \( M \) and \( E \) modes respectively. The quantization procedure can be then carried out as usual with frequencies,

\[ \omega_n = \pm (2n + 3)^{1/2} a^{1/2}, \quad n = 0, 1, 2, \ldots \]

But there is spurious \( \mathrm{cm} \), to remove which we follow the method of Khadkikar and Gupta (1983). For this one must construct \( A \)-gluon states with no quantas in \( \mathrm{cm} \) motion. Still there is the null point motion and its average is deducted from \( \omega_n^2 \) for each gluon. Thus the intrinsic frequencies (energies) are given by (\( \hbar = c = 1 \)):

\[ e_n = \left( 2n + 3 - \frac{3}{A} \right)^{1/2} a^{1/2}. \] (7)

Thus the masses of glue-balls can be calculated using (7) and the properly constructed intrinsic states. These are tabulated (table 3). For comparison the experimental (Chanowitz 1982) candidates are also shown. We also show the results for trigluon glue-balls (Khadkikar and Vinodkumar 1984). These are predictions and in all only two glue-ball candidates are known. It still remains to see the effect of residual interactions as gluons couple to themselves.

It is clear that the mere act of confinement is able to give plenty of information of the strongly interacting particles and one must look for applications in all aspects. We shall present some of these in §4.

### Table 3. Masses of glue-ball states with their possible angular momentum \((J)\) and parity \((P)\) charge conjugation. Labels are given for two- and three-gluon \((EM)\) combinations along with masses of two experimental candidates.

<table>
<thead>
<tr>
<th>Modes</th>
<th>Calculations</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( J^{PC} )</td>
<td>Mass MeV</td>
</tr>
<tr>
<td>EE</td>
<td>0^+, 2^+</td>
<td>1137</td>
</tr>
<tr>
<td>EM</td>
<td>0^-, 2^-</td>
<td>1440</td>
</tr>
<tr>
<td>MM</td>
<td>0^+, 2^+</td>
<td>1703</td>
</tr>
<tr>
<td>EEE</td>
<td>0^+, 2^+, 1^-, 3^-</td>
<td>1971</td>
</tr>
<tr>
<td>EEM</td>
<td>0^+, 2^+, 1^-, 3^-</td>
<td>2296</td>
</tr>
<tr>
<td>EMM</td>
<td>0^+, 2^+, 1^-, 3^-</td>
<td>2489</td>
</tr>
<tr>
<td>MMM</td>
<td>0^+, 2^+, 1^-, 3^-</td>
<td>2720</td>
</tr>
</tbody>
</table>
4. Some applications of RHM

The relativistic wavefunctions in RHM can be transformed to non-relativistic ones with corresponding transformed operators using Foldy-Wouthuysen transformation. Hence all phenomenology with constituent quarks remains essentially valid. This was especially verified in the case of baryonic magnetic moments.

The proton-polarisability calculations eliminating spurious cm and RHM for quarks (Kuyucak 1984) has given much better agreement with experimental parameters than bag models or the non-relativistic oscillator model which over estimates by a factor of 3/2. Moreover, RHM is the only model which has internal consistency in that it gives \( \left( \langle r^2 \rangle_{ch} \right)^{1/2} \) and polarisability sum rule \((0.91 f, 12.7 \times 10^{-53} \text{ cm}^3)\) in agreement with the experimental values \((0.88 f, 14.2 \pm 0.3 \times 10^{-43} \text{ cm}^3)\).

The proton-antiproton annihilation branching ratios to various mesons represent yet another application of the confinement model. While it was found that RHM is superior to other similar models in 3-meson decays, the 2-meson decays representing annihilation of a \( q\bar{q} \) pair to vacuum, presents some uncertainty (Furui et al 1984). The calculated branching ratios in some channels, however, are in remarkable agreement with experimental ones (table 4).

Several more applications such as effective quarkpion coupling, pion-nucleon coupling, nucleon-nucleon force and exotics such as hybrid meiktons using RHM and the gluon confinement scheme as given here are under way.

5. Outlook

It appears that the traditional picture of point hadrons with mutual couplings has been replaced by a few gluon-quark degrees of freedom. This, however, throws a new challenge to understand the nucleus from the ‘quark’ point of view. The conventional independent particle model has to be explained. However, the limitations of independent particle/pair picture have been quite obvious, in the decade long investigations in nuclear matter. Perhaps exchange forces of a molecular type and a geometric model point of view expressed (e.g. see Ramanna and Raghavacharyulu 1975 and references therein) represent premonition of the radical change in understanding of the nuclear structure which is yet so incomplete on a fundamental level.
References

Bogoliubov P N 1972 *Sov. J. Part. Nucl.* 3 71
Isgur N and Carl G 1978 *Phys. Lett.* B74 353
Isgur N and Carl G 1979a *Phys. Rev.* D18 4187
Isgur N and Carl G 1979b *Phys. Rev.* D20 1191
Kuyucak S 1984 *Nucleon polarisability in “The Quark models”* Preprint
Lee T D 1981 *Particle physics and introduction to field theory* (New York: Harwood Academy)
Ramanna R and Raghavacharyulu I V V 1975 *The general many body problem* (eds) R Pratap and J C Parikh
   p. 156
Isobaric degrees of freedom in nuclei as determined from nonrelativistic quark model

C S WARKE
Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

Abstract. Isobaric degrees of freedom $\Delta \Delta$ in nuclei are determined from the quark cluster model of a nucleus. These additional degrees of freedom are brought in by the coloured quark exchange between different nucleon clusters present in nuclei. They are found to be important in the region of momentum transfer near $3.5 \text{fm}^{-1}$. The mass dependence of these isobaric degrees of freedom in nuclei turns out to be $A^{5/6}$.

Keywords. Quark model; nuclear quark cluster model; closed shell nuclei; baryon spectra.

PACS No. 12.35; 12.40; 21.65

1. Introduction

The quark models based on predictions of quantum chromodynamics theory have been quite successful in explaining a large body of data on static properties of hadrons (Chodos et al 1974; De Rujula et al 1975; Dalitz 1977). These models of strong nuclear interactions provide a natural framework for understanding some aspects of nuclear physics which were either ad hoc or were based on flimsy grounds, examples of which are nuclear short distance repulsion, the high momentum components of the elastic and inelastic form factors of nuclei and the isobaric components of the nuclear wave function. Recently Jaffe et al 1984; Close et al 1983 have argued that European Muon Collaboration (EMC) effect may be due to the nuclear environmental effect on the quark structure of a nucleon. They give the mass dependence of this effect. Some of the problems have been tackled in the recent past (Matveev and Sorba 1977; Hagaasen et al 1980; Smirnov and Tchuvilsky 1978; Warke et al 1981) in the six quark system within the framework of different models. In this paper we use the nonrelativistic quark model to describe nuclei as clusters of three-quark baryons. We further assume that the quark structure of nucleons in a nucleus does not change. However, the colour quark exchange takes place between different nucleon clusters in a nucleus. This quark exchange generates the short range nucleon-nucleon correlations in the nucleus, which is vital for any inter-nucleon contribution to the large momentum transfer phenomenon or short-range effects in a nucleus. This contribution is over and above that of the individual nucleons alone. Based on this picture, we develop here a formalism for the isobaric and hidden colour degrees of freedom in nuclei. We study its mass number dependence and correlate it to the EMC effect and its mass dependence.

2. Nuclear quark model and an isobaric component of nuclear wavefunction

Let us consider a nucleus of mass number $A$ and atomic number $Z$ whose number of constituent quarks are $3A$. We assume that baryon structure does not change
appreciably inside the nucleus. As observed in our earlier studies (Shanker and Warke 1980a, b; Warke and Shanker 1979, 1980) the short range correlation effects between nucleons in a nucleus arise naturally from the quark exchange mechanism between different nucleon clusters. The \( n \)th nucleon 3-quark cluster wave function is

\[
\psi^{(n)}(\frac{1}{2}m, \frac{1}{2}t, x) = Q^{(n)}(3n-2, 3n-1, 3n)\chi^{(n)}_{\frac{1}{2}m, \frac{1}{2}t} \times (3n-2, 3n-1, 3n)\psi^{(n)}_{\frac{1}{2}t} (R_n). \tag{1}
\]

In (1), \( \chi^{(n)} \) is the 3-quark spin-isospin wavefunction and \( Q^{(n)} \) is their translationally invariant, correlated space wavefunction which is determined from baryon spectroscopic studies (Shanker and Warke 1980). The centre of mass coordinate of 3-quarks of the \( n \)th cluster is \( R_n \) and \( \psi^{(n)}_{\frac{1}{2}t} \) is the \( n \)th cluster shell model wave function. Consistent with the nuclear shell model, the cluster wave function is

\[
\psi^{(n)}(jm, \frac{1}{2}t, x) = Q^{(n)} \sum_{\mu} (l, m - \mu; \frac{1}{2} \mu^{(n)} | jm) \psi^{(n)}_{\frac{1}{2}m, \frac{1}{2}t} \psi^{(n)}_{\frac{1}{2}t}. \tag{2}
\]

The Pauli principle further restricts the 3A quark wave function to the determinantal form (Warke 1984),

\[
\Psi(1, 2, 3, \ldots; 3A) = \frac{1}{(4!)^{3/2}} \det \left\{ \psi^{(1)}_{\mu_1}, \psi^{(2)}_{\mu_2}, \ldots, \psi^{(A)}_{\mu_A} \right\}. \tag{3}
\]

The wave function in (3) is clearly antisymmetric with respect to the interchange of nucleon clusters and the exchange of quarks within the same nucleon cluster, but is not antisymmetric with respect to the interchange of quarks between different clusters. In order to make it completely antisymmetric in all the 3A quarks, the antisymmetrizer operator \( s \) is applied \( \Psi^A = Cs\Psi \),

\[
s = \left[ 1 + \sum (-1)^s P_s \right]/N^{1/2}.
\]

The phase \( (-1)^s \) is the signature of the permutation \( P_s \). The normalization factor \( N = (3A)!/[4! (3!)^3] \). We observe that \( \Psi \) in (3) is normalized to unity but \( s\Psi \) is not. Therefore the normalization constant \( C \) is introduced in \( \Psi^A \). In writing (3), we tacitly assumed that nuclei under consideration are completely filled shell nuclei and that the cluster wave functions \( \psi_{\mu_i} \) are orthogonal. We further restrict our study to nuclei having \( N = Z = A/2 \). In principle, the nuclear shell model wave function should be obtained from the nuclear interaction as derived from the same quark-quark interaction as that used in the determination of the 3-quark cluster wave function. However, since such an approach gives only the short range repulsion of the nucleon-nucleon interaction correctly, this approach cannot be followed at this stage of our understanding. Therefore, the effect of long-range interaction is included phenomenologically by assuming the nuclear independent particle shell model of a nucleus. The oscillator parameter of this wave function is fixed from the observed nuclear root mean square radius.

As each cluster wave function is a colour singlet, the 3A quark wavefunction \( \Psi \) is also colour singlet. Since \( N = Z \) and we are considering doubly closed shell nuclei, the spin and isospin of the 3A quarks in \( \Psi \) are zero. The quark exchange operator is scalar under colour, spin and isospin transformations. Therefore, \( \Psi^A \) is also a colour singlet and has zero total spin and isospin. One can easily construct the following spin, isospin and
colour exchange operators,
\[ P^\sigma(ij) = \frac{1}{2}(1 + \sigma_i \cdot \sigma_j); \quad P^t(ij) = \frac{1}{2}(1 + t_i \cdot t_j); \quad P^c(ij) = \frac{1}{2} + 2F_i \cdot F_j. \]

In the colour exchange operator \( P^c, F \) are the eight colour SU(3) generators. Through the antisymmetrization process, the spin-isospin exchange operators introduce the \( \Delta(3/2, 3/2) \) spin-isospin components in \( \Psi^A \). Similarly, \( P^t \) introduces the colour octet components (hidden colour) in \( \Psi^A \). As \( \Psi^A \) is spin-isospin and a colour singlet for closed shell nuclei, this mixing occurs in \( \Delta \Delta \) and colour \( 8 \times 8 \) pairs to lowest order. In the following we derive the expressions for their mixing amplitudes.

The \( \Delta \Delta \) wave function can be defined analogous to that for the nuclear wave function in (3). Any two clusters in \( \Psi \) are changed to two delta clusters due to the quark exchange. Let us assume, they occupy the orbitals \( \delta_1 \) and \( \delta_2 \) in place of two nucleons in orbits \( \beta_1 \) and \( \beta_2 \) occupied in \( \Psi \).

\[ \Psi_D = \frac{1}{(A!)^{1/2}} \text{Det} \left\{ \psi^{(1)}_{\delta_1}, \psi^{(2)}_{\delta_2}, \ldots, \psi^{(A)}_{\beta_2} \right\}. \]  

(4)

The properly antisymmetrized \( 3A \) quark wave function is \( \Psi_D = s \Psi_D \). Following the definitions used in nuclear physics, we define the spectroscopic factor \( S \) of a \( \Delta \Delta \) component and \( NN \) component,

\[ S(\Delta \Delta, \delta_1 \delta_2; \beta_1 \beta_2) = |\langle s \Psi_D | \Psi^A \rangle|^2 = N|\langle \Psi_D | \Psi^A \rangle|^2; \]

\[ S(NN) = |\langle s \Psi | \Psi^A \rangle|^2 = N|\langle \Psi | \Psi^A \rangle|^2. \]  

(5)

Note that although \( \Psi_D \) does not have good spin and isospin, it is a colour singlet. The required spin and isospin of \( \Psi_D \) is correctly projected out in the overlap amplitudes defined in (5). Therefore, the complication of constructing good spin and isospin manybody state \( \Psi_D \) is avoided. In the overlap integrals in (5), antisymmetrization introduces quark exchanges between \( A \) clusters. It is expected that quark exchange probability will fall faster with the number of clusters between which the quarks have to be exchanged. This is because quark exchange between different clusters require these clusters to be overlapping. The probability of three nucleons overlapping in a nucleus is smaller than that of two nucleons overlapping each other. We keep the contribution of quark exchanges between two clusters and neglect that coming from the multi-cluster quark exchanges in evaluating the spectroscopic factor in (5). With this assumption we have

\[ N^{1/2} \langle \Psi_D | \Psi^A \rangle \simeq \langle \Psi_D | \Psi_D \rangle - \sum_{m < n} \sum_{i(m), j(n)} P_{i(m), j(n)} |\Psi_D \rangle. \]  

(6)

The sum over \( m, n \) in the second term in (6) is over the \( A \) nucleon clusters in a nucleus. The additional sum \( i(m) \) and \( j(n) \) over the quarks, the \( i \)th quark in the \( m \)th cluster is exchanged with the \( j \)th quark in the \( n \)th cluster \( j(n) \). Because of the spin-isospin \( (3/2, 3/2) \) wave function orthogonality of delta to that of a nucleon \( (\frac{1}{2}, \frac{1}{2}) \), the first term in (6) vanishes. Again from the orthogonality of the shell model single particle wave functions, the second term reduces to

\[ (N)^{1/2} \langle \Psi_D | \Psi^A \rangle = -9 \langle \psi^{(1)}_{\delta_1} \psi^{(2)}_{\delta_2} | P_{i(1), j(2)} | \psi^{(1)}_{\beta_1} \psi^{(2)}_{\beta_2} - \psi^{(2)}_{\beta_1} \psi^{(1)}_{\beta_2} \rangle. \]  

(7)

Now onwards the cluster indices 1 and 2 in (7) can be dropped since there is no possible ambiguity. It is to be noted that the first term in (6) will be unity for \( S(NN) \) calculation.
Let us evaluate separately the colour exchange, spin-isospin exchange and radial exchange matrix elements \( (me) \) in (7). The colour singlet wave functions of two nucleons are \( \xi_c(1, 2, 3) \) and \( \xi_c(4, 5, 6) \) and the colour exchange operator is \( P_c \). Using the determinantal form of \( \xi_c \) the colour part of the \( me \) can be easily calculated

\[
\langle \xi_c(1, 2, 3) \xi_c(4, 5, 6) | P_c(16) | \xi_c(1, 2, 3) \xi_c(4, 5, 6) \rangle = \frac{1}{3}. \tag{8}
\]

The integration over quark internal coordinates in a cluster with the quark space exchange operator in (7) requires the delta space wave function. We assume that spatial quark structure of the delta and a nucleon is the same. Using the baryon wave function derived by Shanker et al 1980a from the fit of nucleon spectroscopic data, one has

\[
\langle Q_\Delta(1, 2, 3) Q_\Delta(4, 5, 6) | P^s(16) | Q_N(1, 2, 3) Q_N(4, 5, 6) \rangle = \frac{27}{5(5)^{1/2}} \exp(-1.87 \beta^2 r^2) = v(r). \tag{9}
\]

In the above equation \( \beta \) is the parameter of the translationally invariant wave function and \( r \) is the relative separation of the centre of masses of the two clusters. The value of \( \beta = 2.02 \text{ fm}^{-1} \) (Shanker et al 1980a). The quark space exchange operator exchanges the position coordinates of quarks 1 and 6 not only in the clusters internal wave functions but also in the relative wave function of their centre of masses. We have neglected it in (9), which is justified because the cluster shell model wave functions are expected to vary slowly with \( r \) relative to the cluster internal wave function. The usual transformation is used to transform the nuclear shell model wave functions in (9) to the \( ls \) coupling scheme. This allows us to express the six quark spin isospin wave functions corresponding to a nucleon or a delta in the following form

\[
| \chi_{N,D}^2 \chi_{N,D}^2 SM_4; \chi_{N,D}^2 \chi_{N,D}^2 TT_3 \rangle.
\]

Although lengthy, it is straightforward to evaluate now the spin isospin \( me \).

\[
\langle \chi_{3/2 3/2}(1, 2, 3) \chi_{3/2 3/2}(4, 5, 6) S' M'_4; T'T'_3 | P^s(16) P^t(16) | \chi_{1/2 1/2}(1, 2, 3) \chi_{1/2 1/2}(4, 5, 6) S M_4 T T_3 \rangle = \frac{4}{81} [(9 - 4S)(9 - 4T)]^{1/2} \delta(SM_4 T T_3; S'M'_4 T'T'_3). \tag{10}
\]

Using the results from (8) to (10) in (7), the six quark \( me \) is reduced to the usual shell model type \( me \) of a two-body interaction for the centre of masses of the clusters. This effective two-body interaction is

\[
v(r, ST) = -\frac{4}{27} [(9 - 4S)(9 - 4T)]^{1/2} v(r). \tag{11}
\]

The exchange integral in (7) is reduced to

\[
\langle N \rangle^{1/2} \langle \Psi_D | \Psi_A \rangle = \sum_{STL} X(StL) \langle \delta_1 \delta_2; L | v_a \beta_1 \beta_2; L \rangle. \tag{12}
\]

The antisymmetric two-body \( me \) has \( v_a = v[1 + (-1)^{S+T+1} P^s(12)] \), and the geometric
transformation factor $X$ is

$$X(STL; \delta\delta'; \beta\beta') = \sum_{jM} (jm, j'm' \mid JM)_\beta (jm, j'm' \mid JM)_\beta (\frac{1}{2}t, \frac{1}{2}t' \mid TT_3)_\beta$$

$$[(\frac{1}{2}t, \frac{1}{2}t' \mid TT_3)_\beta (2L + 1)(2S + 1)((2j + 1)(2j' + 1))]^{1/2}$$

$$[(2j + 1)(2j' + 1))]^{1/2}$$

$$(l \quad \frac{1}{2} \quad J) \quad (l' \quad \frac{1}{2} \quad J')$$

$$(L \quad S \quad J)_{\beta} \quad (L \quad S \quad J)_{\delta}$$

In (13) the subscript $\beta$ and $\delta$ indicate the $NN$ pair in shell model orbits $\beta = (n_p l_p j_p m_p)$ and similarly for $\beta'$ and a pair of deltas. The corresponding values of quantum numbers are to be used in (13). Equation (12) gives the probability amplitude for finding the two 3-quark clusters of nucleons in a nuclear shell model states $\beta$ and $\beta'$ and in the two deltas in shell model states $\delta_1$ and $\delta_2$. The probability $P(\delta\delta'; \beta\beta')$ is a square of this amplitude. The total probability of finding a pair of $\Delta\Delta$ isobars in a nucleus would be the sum over the occupied nucleon orbitals $\beta\beta'$ in a given nucleus and over all the possible shell model states $\delta\delta'$ of the isobar pair (including their charge states). The sum over $\delta\delta'$ is performed using the orthogonality relation of the $9-j$ symbols and the completeness relation of the harmonic oscillator states. The sum is then expressed in terms of the probability of creating an isobar pair $\Delta\Delta$ at positions $r_1$ and $r_2$ inside a nucleus. As we are studying only closed shell nuclei with $N = Z$, the sum over $m, m'$ and $t, t'$ ($\beta, \beta'$) in $P_D$ can be easily performed. The isospin state sums of the $\Delta\Delta$ pair further simplify this expression. The reduced form of $P_D$ of creating a $\Delta\Delta$ pair at positions $r_1$ and $r_2$ inside the nucleus becomes

$$P_D(r_1, r_2) = \sum_{nlj, n'l'j'} (2T + 1)(2L + 1)(2S + 1)(2J + 1)(2j + 1)(2j' + 1)$$

$$\left\{ \begin{array}{c|c|c} l & 1/2 & j \\ l' & 1/2 & j' \\ L & S & J \end{array} \right\}_J^{2} \left| \left\langle r_1, r_2 \mid v_a(ST) \right\rangle |nl, n'l' ; LM \right\rangle \right|^2.$$ (14)

The sum over $nlj, n'l'j'$ in (14) extends over the occupied orbitals in the nuclear shell model wave functions. Let us transform the $\Delta\Delta$ coordinates in the centre of mass and relative coordinates $R = (r_1 + r_2)/2$ and $r = r_1 - r_2$ and use Moshinsky transformation brackets $\langle nl, n'l' \parallel L \parallel nl, NA \rangle$. After integrating $P_D$ over the coordinate $R$ and angular orientation of $r$, one obtains the probability of finding an isobar pair $\Delta\Delta$ in a nucleus at a relative separation $r$ between them.

$$\left| \left\langle r \mid v_a |nl, n'l' ; L \right\rangle \right|^2 = \sum_{nl, l_1, NA} \left[ 1 + (-1)^{S+T+1+1} \right]$$

$$\langle nl, n'l' \parallel L \parallel n_1 l_1, NA \rangle^2 Q^2_{nl_1}(r) v^2(r, ST).$$ (15)

In (15) $Q_{nl}(r)$ is the radial wave function of the relative motion of the $\Delta\Delta$ pair inside the nucleus. From (14) and (15) we obtain the final expression for finding an isobar pair in a nucleus at separation $r$

$$P_3(r) = \sum_{nST} (2T + 1)[1 + (-1)^{S+T+1+1}] C_{ST}^2 v^2(r, ST) Q_{nl}(r).$$ (16)
The nuclear structure dependent constants $C_{\lambda}^{S_T}$ are defined as the sum over occupied orbitals $a, b$ in a nucleus

$$C_{\lambda}^{S_T} = \sum_a \sum_{\lambda} (2\lambda + 1)(2S + 1)(2J_a + 1)(2J_b + 1) \begin{pmatrix} l_a & 1/2 & j_a \\ l_b & 1/2 & j_b \\ \lambda & S & J \end{pmatrix}^2$$

$$\times \sum_{NL} \langle nl, NL | \lambda | n_a l_a, n_b l_b \rangle^2. \quad (17)$$

3. Delta isobar components in $^{16}$O and $^{40}$Ca nuclei

In these doubly closed shell nuclei, the sum over spin and isospin can easily be carried out. The probability $P_D(r)$ in (16) becomes

$$P_D(r, ^{16}$O) = v^2 [82.95Q_0^2 + 5.93Q_{10}^2 + 29.62Q_{02}^2 + 80.6Q_{01}^2];$$

$$P_D(r, ^{40}$Ca) = v^2 [274.1Q_0^2 + 51.9Q_{10}^2 + 4.5Q_{20}^2 + 259.5Q_{02}^2 + 22.5Q_{12}^2 + 403.0Q_{21}^2 + 60.45Q_{31}^2], \quad (18)$$

where $v(r) = 27/5(5)^{1/2} \exp (-1.8 \beta^2 r^2)$; $\beta = 2.02 \text{ fm}^{-1}$. In (18) $Q_{nl}(r)$ are the relative wave functions of a pair of nucleons in the nucleus. Before we proceed further, a comment is in order about the use of the term ‘probability’. In this paper as we are only studying the isobar degrees of freedom and are not calculating the nucleon and hidden colour components, we cannot normalize the total probability to unity. In this sense the use of this term is improper. Therefore, from now on, we will simply use the term ‘isobar degrees of freedom’ instead of ‘probability’ for finding an isobar pair in a nucleus.

We notice that due to a factor $v^2(r)$, $P_D(r)$ in (18) is maximum at short distances for which $(3.6)^{1/2} \beta r = 1.9, \beta r \approx 1$ or $r \approx 0.26 \text{ fm}$. At such short inter-nucleon distances in a nucleus, the nucleon three-quark clusters overlap and introduce additional isobar and hidden colour degrees of freedom in the nuclear wave function. This corresponds to a region of momentum transfer near $3.5 \text{ fm}^{-1}$. At such short inter-particle distances only the $l = 0$ wave functions contribute to $P_D(r)$ since $Q_{nl}(r)$ approaches zero for $l \neq 0$ and for $l = 0$, $Q_{n0}(n \neq 0)$ is smaller than $Q_{00}$ as $r \to 0$. Thus it is clear that the isobar components of nuclear wave functions are important only at small inter-nucleon separations. Similar conclusions are arrived at for the hidden colour degrees of freedom in a nucleus. It is expected that at even shorter distances than this, the dominant part of the nuclear wave function will be the hidden colour component (Warke et al 1981). These are the additional degrees of freedom in nuclei besides those of nucleons which are brought in by the quark structure of the nucleon. Some experimental indications support the observation of such additional degrees of freedom in nuclei. The understanding of the discovery of Gamow-Teller resonances in $(n, p)$ scattering off nuclei requires the presence of virtual deltas in a nucleus (Goodman 1984). It is expected that there could be a significant enhancement of the virtual pion field inside a nucleus in the large momentum transfer processes $\approx 2.5 \text{ fm}^{-1}$, close to the distance where $P_D(r)$ is maximum. It leads to the prediction of an enhancement of non-strange components of the nuclear sea, which can be made to agree with the EMC data at small $x$. Here we are not concerned with the quantitative prediction of our theory. Instead, we would like to
qualitatively study how this $\Delta A$ component of the nuclear wave function varies with the nuclear mass number $A$ for doubly closed shell nuclei.

As discussed above, we neglect the non-zero angular momentum contribution and the $n \neq 0$ contribution to the total integrated value of $P_D$,

$$P_D(^{16}\text{O}) = 82.95[27/5(5)^{1/2}]^2/(3.6\beta^2 b^2 + 1)^{3/2};$$
$$P_D(^{40}\text{Ca}) = 274.1[27/5(5)^{1/2}]^2/(3.6\beta^2 b^2 + 1)^{3/2}. \quad (19)$$

We take nuclear shell model value of $b = 1.01 A^{1/6}$ fm. The study of the coefficients of $Q_{00}^2$ in (18) for different nuclei reveals that they are approximately proportional to $A^{4/3}$ with the proportionality constant 2.05. Using this observation, one can express $P_D$ in (19) as a function of $A$ as follows,

$$P_D(A) \approx 0.2 A^{5/6}. \quad (20)$$

The total isobar degrees of freedom in a nucleus arising from the nucleon quark structure increase with the mass number $A$ as the $5/6$th power of $A$. The additional isobar degrees of freedom per nucleon in a nucleus would be $0.2 A^{-1/6}$. If we associate these additional degrees of freedom to the slight increase in size of a nucleon in the nucleus, our result would predict that this increase in nucleon size inside a nucleus would become smaller for heavier nuclei.

4. Conclusions

The quark cluster model of a nucleus is developed and is used to study the isobaric $\Delta A$ degrees of freedom in nuclei. The internal quark cluster wave function is taken from the baryon spectroscopic data analysis. These additional degrees of freedom are found to be important at short inter-nucleon separation phenomenon in a nucleus corresponding to the region of momentum transfer near $3.5 \text{ fm}^{-1}$. They are brought in by the quark structure of nucleons and are found to increase with the nuclear mass number $A^{5/6}$. If one associates with this effect an equivalent increase in size of a nucleon in the nuclear medium, our result would predict this increase of nucleon size to depend on the mass number. It decreases in an exponential form $A^{-1/6}$ as one approaches heavier nuclei. Contrary to this result, most predictions are linear in the effective density of the nucleus (Arnold et al 1984; Thomas 1984).

References

Hogaasen H et al 1980 Z. Phys. C4 131
Shanker R and Warke C S 1980b Hadronic J. 3 1108
Thomas A W 1984 University of Adelaide Preprint ADP 300
Warke C S 1984 Pramana 22 247
The fifth interaction: universal long range force between spins

T PRADHAN, R P MALIK and P C NAIK*
Institute of Physics, Sachivalaya Marg, Bhubaneswar 751 005, India
* Department of Physics, S.C.S. College, Puri 752 001, India

Abstract. In this paper we present a review of our investigations on universal long range force between spins mediated by a massless axial vector gauge field which we name as “axial photon”. The invariance of the Lagrangian field theory of particles, possessing spin degrees of freedom, under local Lorentz transformations, necessitates the introduction of such an axial vector gauge field which interacts with spin current of the particles. Classical as well as quantum dynamics of electrons interacting with photon and axial photon are worked out. The new interaction is found to be asymptotically free. It is shown that QED can be made finite if the coupling strengths of electron to photon and axial photon can be made equal. Experimental consequences of the existence of axial photon are discussed and the strength of the interaction is estimated by comparing predictions of the theory with experiments.

Keywords. Fifth interaction; axial photon; local Lorentz group; divergences; stability of the electron; asymptotic freedom; hyperfine anomaly; phase of the wavefunction.

PACS No. 12-90

1. Introduction

There are at present four kinds of known fundamental interactions in nature namely: strong, electromagnetic, weak and gravitational. It is the present-day conviction that all these interactions stem from certain symmetries and are mediated by gauge fields that become necessary to restore symmetry when the parameters of global transformation are made space-time dependent. For instance the electromagnetic field is required to restore phase invariance of the Lagrangian when the parameter of the phase transformation is made space-time dependent. The coloured gluons (Fritzsch et al 1973), the $W$ and $Z$ bosons (Glashow 1980; Weinberg 1980; Salam 1980) that mediate strong and weak interactions are gauge particles associated with space-time dependent $SU(3)_{C}$ and $SU(2) \times U(1)$ transformations respectively. The gravitational interaction is thought to be mediated by spin-2 massless gauge particle required in space-time dependent translation group, (Cho 1976; Hayward 1979). It is natural to speculate about gauge particle associated with the Lorentz group. Although the latter has been considered by various authors (Utiyama 1956; Kibble 1961; Sciama 1962) who obtain the gravitational interaction by gauging the Lorentz group it is somewhat artificial. In all gauge theories the gauge particle usually couples to the conserved current of the global group which is gauged. In Einstein’s theory the gravitational field couples to energy-momentum tensor (Carmeli 1982) which is the conserved current for the space-time translation group. It is therefore natural to associate gravitational interaction with the local translation group. In the Lorentz group the conserved current is the angular momentum tensor density and one would therefore expect that the gauge particle in the local Lorentz group should couple to this density. In other words, a local Lorentz group
should lead to interaction between spins mediated by an axial vector gauge boson. We have shown that this is indeed the case (Naik and Pradhan 1981). Our approach to gauging the Lorentz group is different from those of Kibble, Sciama and Utiyama. We name the gauge particle as "axial photon". When the gauge theory is worked out in detail one finds that axial photon couples to the axial vector currents of the Dirac particle as well as that of the photon and the space integrals of the space components of these currents are the spin vectors of the respective particles. We thus have a fifth interaction which operates universally between spins of particles. Further we have a mechanism of dynamically measuring spin through axial photon probe.

That such an interaction should exist, can be inferred from physical considerations without going to gauge theory. This had been done when axial photon was first introduced (Pradhan and Lahiri 1974). Their motivation for introducing the axial photon was to give stability to the classical electron and also to make quantum electrodynamics divergence free. It was argued that if an axial vector boson is coupled to the axial vector current of the electron it would result in spin-spin interaction which is attractive for parallel spins. The parts of classical spinning electron would therefore attract each other and compensate the force of Coulomb repulsion. It turns out, as we shall see in § 3 that this is indeed the case but only when the spin-spin coupling strength is equal to the strength of the Coulomb interaction. It also turns out that in quantum theory the divergence associated with electron-self-energy due to exchange of virtual photon cancels with that arising from exchange of virtual axial photon under exactly the same condition (Pradhan and Lahiri 1974). However, the axial photon coupling to electron and photons with such large strength would destroy the excellent agreement of present quantum electrodynamics with experiment unless of course the axial photon acquires a large mass through spontaneous breaking of the gauge symmetry. This possibility was considered by Pradhan and Lahiri. The large mass suppresses physical processes involving axial photon exchanges, but do not affect divergent diagrams.

As mentioned earlier, the force arising out of exchange of axial photon between particles with spin is attractive when the spins are parallel and repulsive when they are antiparallel. Strikingly enough, this is exactly the situation with the experiments carried out with polarized laser beams by Tam and Happer (1977) who found that two right circularly polarized laser beams passing through dense sodium vapour attract each other while those with opposite circular polarization repel. They also found that the force between the beams was of long range. The sodium vapour acts as a spin polarizable medium which on account of asymptotic freedom of axial photon coupling enhances the spin-spin force between laser beams. By analysing the experimental results of Tam and Happer and considering the enhancement by spin-polarizability of the medium, Naik and Pradhan found that good agreement between theory and the experimental results, over the range of densities of sodium vapour used in these experiments, can be achieved if the coupling strength is taken to be \( \alpha_g = g^2/4\pi = 0.7 \times 10^{-9} \).

An independent estimate of the coupling strength can be made from the contribution of the long range force between nuclear and electronic spin to hyperfine splitting in the hydrogen and the deuterium atoms for which only a small discrepancy exists between theory and experiment. This discrepancy can be accounted for by the long range spin-spin interaction if the axial photon coupling strength is taken to be \( \alpha_g = g^2/4\pi = 10^{-13} \) which is several orders of magnitude smaller than that estimated from the force between circularly polarized laser beams in sodium vapour. We take the latter value to
be more reliable since precision of measurement of hyperfine splitting is much higher than that of the laser beam experiment.

In view of these two experimental evidences for the existence of the fifth interaction we think it worthwhile to review the classical and quantum aspects of this interaction. In this paper we begin with a presentation of the gauging of the Lorentz group in § 2. The classical dynamics of the theory is worked out in § 3. Equations of motion for the axial photon field and its spin source are worked out and their static solution for point sources obtained. Stability of the classical electron and radiation of axial photon by spin sources are discussed in the classical framework. Section 4 is devoted to the quantum dynamics of the theory where questions such as the removal of divergences, asymptotic freedom of the interactions, its contribution to hyperfine splitting and two-body bound states resulting from long range spin-spin forces are discussed. In this section we also discuss a Bohm-Aharonov type of experiment which can establish the existence of the fifth interaction as well as measure its strength.

2. Local Lorentz group

Invariance of Lagrangian density under global Lorentz group of transformations:

\[ x_\mu \rightarrow x'_\mu = x_\mu + \alpha_\mu \nu x_\nu \]

\[ \alpha_\mu \nu = -\alpha_\nu \mu \]  

(1)

leads to conservation of angular momentum \( J_\mu \nu \):

\[ \frac{\partial J_\mu \nu}{\partial t} = 0; \quad J_\mu \nu = \int d^4x \sigma_\lambda M_{\mu, \nu \lambda} \]

\[ \partial_\lambda M_{\mu, \nu \lambda} = 0 \]  

(2)

where \( M_{\mu, \nu \lambda} \) is the angular momentum density. As mentioned in § 1 one expects that in the local Lorentz invariant theory the gauge-field would couple to the conserved current \( M_{\mu, \nu \lambda} \). This implies that the gauge field would be a third rank tensor \( B_{\mu, \nu \lambda} \). That this is actually the case can be seen as follows:

\[ \mathcal{L}^{(0)} = \frac{1}{2} [\tilde{\psi} \gamma_\mu \partial_\mu \psi - (\partial_\mu \tilde{\psi}) \gamma_\mu \psi] - m \tilde{\psi} \psi. \]  

(3)

The above Dirac Lagrangian density which is invariant under global transformation (1) under which,

\[ \psi (x) \rightarrow \psi' (x) = \exp \left[ \frac{i}{2} \sum_\nu \alpha_\nu \lambda \right] \psi (x) = \Omega \psi (x) \]

\[ \sum_\nu \lambda = \frac{1}{4} (\gamma_\nu \gamma_\lambda - \gamma_\nu \gamma_\lambda) \]  

(4)

ceases to be so when the parameters \( \alpha_\mu \nu \) are made space-time dependent. However, the Lagrangian

\[ \mathcal{L}_D = \frac{1}{2} [\tilde{\psi} \gamma_\mu D_\mu \psi - (D_\mu \tilde{\psi}) \gamma_\mu \psi] - m \tilde{\psi} \psi, \]  

(5)

where \( D_\mu \psi = \partial_\mu \psi + ig B_\mu \psi \),

and \( B_\mu = B_{\mu, \nu \lambda} \sum_\nu \lambda, \)
remains invariant when the massless gauge fields $B_\mu$ transform as:

$$B_\mu(x) \to B'_\mu(x') = \Omega(x)B_\mu(x)\Omega^{-1}(x) - \frac{1}{ig}(\partial_\mu\Omega(x))\Omega^{-1}(x).$$  (7)

At this stage it is necessary to point out that in the above, we have considered the invariance of the Lagrangian density which for global transformation ensures the invariance of the action $S = \int d^4x L(x)$. For local transformation this is not the case; one needs extra conditions because,

$$S = \int d^4x L(x) \to S' = \int d^4x' L'(x') = \int d^4x J L''(x'),$$

where $J$ is the Jacobian for the transformation of the four-volume element. For local transformation this is not the case.

The latter is automatically satisfied for global transformation. For our local Lorentz group this is equivalent to,

$$\partial_\mu \sigma^{(x)}_\mu = 0.$$  (10)

In the Kibble-Sciama (Kibble 1961; Sciama 1962) approach to local Lorentz group this restriction is not put, they use vierbeines to define covariant derivatives and their theory leads to Einstein-Cartan theory of gravitation.

The tensor gauge field $B_{\mu,\nu,\lambda}$ is antisymmetric in the last two indices. We can decompose it into symmetric and anti-symmetric parts in the first and last two indices as:

$$B_{\mu,\nu,\lambda} = \frac{1}{2} \varepsilon_{\mu\nu\lambda\zeta} a_\zeta + S_{\mu,\nu,\lambda},$$

where $a_\mu$ is an axial vector field and,

$$S_{\mu,\nu,\lambda} = \frac{1}{2} (B_{\mu,\nu,\lambda} + B_{\nu,\mu,\lambda} + B_{\lambda,\nu,\mu})$$

is a symmetric tensor field which on account of (10) transform as,

$$a_\mu(x) \to a'_\mu(x') = a_\mu(x) + 2/3 \alpha_{\mu\nu}(x)a_\nu(x) - \frac{1}{g} \partial_\mu\Lambda$$

where

$$\partial_\mu \sigma^{(x)}_{\mu\nu\lambda} = \varepsilon_{\mu\lambda\rho} \partial_\rho \Lambda$$

and

$$S^{(x)}_{\mu,\nu,\lambda} \to S'^{(x)}_{\mu,\nu,\lambda} = \Omega(x)S^{(x)}_{\mu,\nu,\lambda}\Omega^{-1}(x).$$

We shall name the axial vector field as “axial photon” field. Substituting (6) and (11) into (5) we obtain

$$L_D = L_D^{(0)} + L_D^{(i)}$$

where

$$L_D^{(i)} = gM_{\mu,\nu,\lambda} B_{\mu,\nu,\lambda}$$

Or

$$L_D^{(i)} = \frac{ig}{4} \varepsilon_{\mu\nu\lambda\zeta} \bar{\psi}(\gamma_{\mu}\sum_{\nu\lambda} + \sum_{\nu\lambda}\gamma_{\mu})\psi a_\zeta$$

$$= 3/2 g\bar{\psi}s^{\mu}_{\nu\lambda}\psi a_\mu$$

(13)
The fifth interaction

which is equivalent to taking

\[ D_\mu \psi = \partial_\mu \psi - 3/2 g \gamma_5 \psi a_\mu. \]  

(14)

It will be noticed that the tensor field \( S_{\mu \nu \lambda} \) does not couple to the Dirac field; only the axial vector field \( a_\mu \) couples to the axial vector current \( \bar{\psi} \gamma_5 \gamma_\mu \psi \) of the Dirac field. The space integral of the space components of this current is the spin of the Dirac particle and that of the time component its helicity. It thus follows that the axial vector field couples to the spin and helicity of the Dirac particle just as the Maxwell field couples to its electric charge and current. Thus axial photon provides a means of dynamical measurement of spin in the same sense as photon measures the charge of the electron.

The kinetic energy term of the gauge field must now be included in the Lagrangian. It is given by

\[ \mathcal{L}_G = -\frac{1}{4} B_{\mu \nu} B^{\mu \nu} \]  

(15)

where

\[ B_{\mu \nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + ig [B_\mu, B_\nu] \]

(16)

The axial photon part of \( \mathcal{L}_G \) is:

\[ \mathcal{L}_G^{\text{axial}} = -\frac{1}{4} (\partial_\mu a_\nu) (\partial_\nu a_\mu) - \frac{3g^2}{16} (a_\mu a_\mu)^2. \]  

(17)

Since the photon has spin it will also couple to axial photon. To obtain the interaction Lagrangian we start with

\[ \mathcal{L}_M = -\frac{1}{4} (D_\mu A_\nu - D_\nu A_\mu) (D_\mu A_\nu - D_\nu A_\mu), \]  

(18)

where

\[ D_\mu A_\nu = \partial_\mu A_\nu + ig B_\nu;_\lambda \sum_{\nu, \lambda} A_\lambda, \]  

(19)

with

\[ \sum_{\nu, \lambda} = -i (\delta_{\nu 2} \delta_{\gamma \nu} - \delta_{\nu \gamma} \delta_{\gamma 2}). \]  

(20)

Substituting (11) and (20) in (19) and evaluating (18) gives the following interaction Lagrangian between photon and axial photon

\[ \mathcal{L}_{MG} = -ge_{\mu \nu \lambda \sigma}(\partial_\mu A_\nu) A_\lambda a_\sigma - \frac{1}{2} g^2 [a_\mu A^2 - (a \cdot A)^2]. \]  

(21)

It will be noticed that this Lagrangian is not invariant under the electromagnetic gauge transformation

\[ A_\mu \to A'_\mu = A_\mu + \frac{1}{e} \partial_\mu \phi. \]  

(22)

This can be rectified by replacing \( A_\mu \) occurring in the Lagrangian by \( (A_\mu + \lambda \partial_\mu \phi) \) where \( \Phi(x) \) is a massless scalar field which transforms as

\[ \phi(x) \to \phi'(x) = \phi(x) - \frac{1}{\lambda e} \alpha(x). \]  

(23)

In order to ensure the invariance of the kinetic energy term \( -\frac{1}{2} (\partial_\mu \phi)(\partial_\mu \phi) \) of the scalar field we add a counter term \( -\frac{1}{2} (\partial_\mu \phi) A_\mu \) to the Lagrangian. In order to decouple the scalar field from the rest, one goes to the limit \( \lambda \to 0 \). This has been discussed in detail (Naik and Pradhan 1981) and is a slightly modified version of Stueckelberg’s (1938) compensating field method.
3. Classical dynamics of axial vector gauge field

3.1 Equations of motion for the field and its source

The Lagrangian density for the field interacting with its spin source as worked out in the preceding section can be written as

\[ \mathcal{L} = \mathcal{L}^{(0)} + \mathcal{L}^{(0)} + \mathcal{L}^{(1)} + \mathcal{L}^{(1)} = \mathcal{L}^{(0)} + \mathcal{L}^{(1)}, \]  

(24a)

where

\[ \mathcal{L}^{(0)} = -\frac{1}{2} (\partial \mu a) \partial \nu a - \frac{1}{4} F_{\mu \nu} F_{\mu \nu} + \frac{1}{2} \left[ \bar{\psi} \gamma_{\mu} \partial_{\nu} \psi - (\partial_{\mu} \bar{\psi}) \gamma_{\mu} \psi \right] - m \bar{\psi} \psi, \]  

(24b)

\[ \mathcal{L}^{(1)} = -\frac{3g^2}{16} (a_{\mu} a_{\mu})^2 - \frac{1}{2} g^2 \left[ a^2 A^2 - (a \cdot A)^2 \right] + s_{\mu}(x) a_{\mu}(x), \]  

(24c)

where

\[ s_{\mu}(x) = \frac{3}{2g} \bar{\psi} \gamma_{\mu} \gamma_{\nu} \psi - g e_{\mu \nu \lambda \xi} (\partial_{\nu} A_{\lambda}) A_{\xi}, \]  

(24d)

is the spin current density of the electron and the photon. If we add a four divergence \( \frac{1}{2} \partial_{\mu} (a_{\nu} \partial_{\nu} a_{\mu}) \) to the \( \mathcal{L}^{(0)} \) in (24b), the kinetic energy of the axial photon can be written as

\[ \mathcal{L}^{(axial)} = -\frac{1}{4} f_{\mu \nu} f_{\mu \nu}, \]  

(25a)

where

\[ f_{\mu \nu} = \partial_{\mu} a_{\nu} - \partial_{\nu} a_{\mu}, \]  

(25b)

and the equations of motion that follow from the Lagrangian are,

\[ \begin{cases} \partial_{\nu} f_{\mu \nu} = s_{\mu}(x) \\ e_{\mu \nu \lambda \xi} \partial_{\nu} f_{\lambda \xi} = 0 \end{cases} \]  

(26)

if we neglect nonlinear terms of order \( g^2 \) in (24c). In terms of \( \mathcal{E}\_i = f_{0i} \) and \( \mathcal{B}_i = \frac{1}{2} e_{ijk} f_{jk} \) these equations read

\[ \begin{cases} \nabla \times \mathcal{E} - \frac{1}{c} \frac{\partial \mathcal{B}}{\partial t} = 0, \\ \nabla \cdot \mathcal{B} = 0 \end{cases} \]  

\[ \begin{cases} \nabla \times \mathcal{B} - \frac{1}{c} \frac{\partial \mathcal{E}}{\partial t} = s(x) \nabla \cdot \mathcal{E} = h(x) \end{cases} \]  

(27)

where \( h(x) = s_{0}(x) \) is the helicity density and together with spin density \( s(x) \) obeys the continuity equation:

\[ \partial_{\mu} s_{\mu}(x) = 0. \]  

(28)

The equation for the axial vector potential \( a_{\mu} \) that follows from (24b) when terms of order \( g^2 \) are neglected, are:

\[ \Box a_{\mu} = s_{\mu}(x). \]  

(29)

This equation can also be regarded as Hamilton’s equations:

\[ \frac{\partial \pi_i}{\partial t} = \delta_{ij} \frac{\delta H}{\delta (\partial_{\mu} a_{i})} - \frac{\delta H}{\delta a_{i}}, \]  

(30)

where

\[ \pi_i = -\frac{\delta H}{\delta \left( \frac{\partial a_{i}}{\partial t} \right)} \]  

(31)
The fifth interaction

is the field canonically conjugate to \( a_i \) and

\[
H = \int d^3x \left\{ \frac{1}{2} (\nabla \times a)^2 + \frac{1}{2} \left( \frac{\partial a}{\partial t} \right)^2 - s \cdot a \right\} + H_0,
\]

\( H_0 \) being the energy of the bare source. It will be seen that the axial vector field couples to the spin of the source.

The equation of motion for the spin of the source can be obtained by using the Poisson Bracket equation:

\[
\frac{\partial S_i}{\partial t} = i [H, S_i].
\]

Using the definition,

\[
S_i = \int d^3x \, s_i(x),
\]

and the ansatz

\[
[s_i(x), s_j(x')]_{t=t} = i \epsilon_{ijk} \delta(x - x') s_k(x),
\]

yields

\[
\frac{\partial S_i}{\partial t} = - \epsilon_{ijk} \int d^3x \, s_j(x,t) a_k(x,t).
\]

This equation shows that the axial photon field exerts a torque on the spin the direction of which is perpendicular to the field potential \( a \) as well as to the spin of the source so that \( S^2 = \text{constant} \). If \( a \) were space-time independent, the solution of (35) would give a gyration of the spin like that of a spinning top in a gravitational field. This situation is to be contrasted with the interaction of the photon field on a charged particle which gets accelerated by it.

3.2 Static field of a spin-source

Solution of (29) can be written as

\[
a_i(x) = \frac{1}{4\pi} \int d^3x' \frac{s_i(x')}{|x - x'|}
\]

For the spin distribution we can take

\[
s_i(x) = g \rho(x) S_i,
\]

where \( \rho(x) \) is the spin distribution function. Substitution of (37) in (36) leads to

\[
a_i(x) = \frac{S_i}{4\pi} \int d^3x' \frac{g \rho(x')}{|x - x'|}
\]

For a point spin source \( \rho(x') = 3/2 \delta(x') \)

\[
a_i(x) = \frac{3g S_i}{8\pi |x|},
\]

and

\[
\mathbf{\xi} = - \nabla \times \mathbf{a} = \frac{3g}{8\pi} \left( \frac{\mathbf{r} \times \mathbf{s}}{r^3} \right)
\]

\[
\mathbf{\zeta} = - \frac{\partial \mathbf{a}}{\partial t} = 0.
\]
The situation here is to be compared with static photon field of a steady electric current which produces a steady magnetic field but no electric field. It will be seen from (40) that for a right circularly polarized light beam travelling in a straight line for which the spin is in the direction of motion, the \( \mathcal{E} \) lines of force will be circles around the beam axis just like magnetic lines of force \( \mathbf{B} \) around a straight wire carrying current. On the other hand for circularly polarized light beam travelling in a circular path, the \( \mathcal{E} \) lines of force would be on the surface of a torus around the path like those of \( \mathbf{B} \) lines around a circular current. It then follows that the \( \mathcal{E} \) lines for an infinitely long helical light path as in a selfoc fibre would be exactly like the \( \mathbf{B} \) lines produced by a solenoid, i.e. the \( \mathcal{E} \) lines would be parallel to the fibre and confined to the inside of the fibre.

3.3 Stability of classical electron

It is well known that classical finite size electron is not stable; its parts would fly apart due to repulsive Coulomb interaction which cannot be prevented unless forces of non-electromagnetic origin are invoked. It will readily be seen that such forces would arise from the new interaction. The static fields produced by electrons' spin and electric charge would be

\[
\mathbf{E} = \frac{e\mathbf{r}}{4\pi r^3}, \quad \mathbf{B} = 0; \quad \mathcal{E} = \frac{3g}{8\pi} \mathbf{r} \times \mathbf{s}; \quad \mathcal{B} = 0. \quad (42)
\]

The \( \mathbf{E} \) and \( \mathcal{E} \) lines of force are shown in figures 1a, b. These fields would give rise to a self-force:

\[
\mathbf{F} = e\mathbf{E} - \frac{3g}{2} \mathbf{s} \times \mathcal{E} = \frac{e^2}{4\pi r^3} \frac{9g^2S^2}{16\pi r^3}, \quad (43)
\]

which can be made to vanish by suitably choosing 'g'

\[
(3/2g)^2 = \frac{e^2}{S^2} \quad \text{or} \quad \frac{9}{4}g^2 = \frac{e^2}{S^2}. \quad (44)
\]

This would ensure stability of the electron. For an extended electron calculations will be somewhat more complex. We thus see that the fifth interaction can provide a "Poincare stress" for ensuring stability of the classical electron.

3.4 Static interaction between two spins

The potential energy between two particles possessing spins, one located at the origin and another at \( \mathbf{r} \) is given by

\[
v_{AB}(\mathbf{r}) = -\int d^3r' \langle a_i^{\dagger}(\mathbf{r}) a_j(\mathbf{r}') \rangle. \quad (45)
\]

Using expression (39) for \( a_i(\mathbf{r}) \) in (45) we obtain after some manipulations,

\[
v_{AB}(\mathbf{r}) = -\frac{(3g/2)^2}{2\pi |\mathbf{r}|} [S_A \cdot \mathbf{S}_B], \quad (46)
\]

which shows that the force between two spins is attractive when the two are parallel while it is repulsive when they are anti-parallel. This can be considered as the Coulomb law for the fifth interaction. It is important to note that identical (parallel) spins attract
while identical charges repel. Therefore for an extended particle with charge and spin, the instability caused by repulsion of different parts of the charge distribution would be opposed by the attraction between parts of spin-distribution. We have already seen this from a slightly different consideration in §3.3.

3.5 Radiation of axial photon

Solution to the wave equation:

\[ \nabla^2 \mathbf{a} - \frac{1}{c^2} \frac{\partial \mathbf{a}}{\partial t} = 4\pi g s(r) \]  

(47)

can be written as

\[ \mathbf{a}(rt) = \int d^3r' \int d't' \frac{s(r',t')}{|r-r'|} \delta \left( t - t' \frac{r-r'}{c} \right). \]  

(48)

For the sinusoidal time variation of the spin density

\[ s(r, t) = s(r) \exp(-i\omega t) \]

equation (48) becomes,

\[ \mathbf{a}(r) = \int d^3r' s(r') \frac{\exp[i\mathbf{k} \cdot (r-r')] |r-r'|}{|r-r'|}. \]  

(49)
With \( k = \omega/c \) and for the limit \( kr \gg 1 \) this reduces to

\[
\lim_{kr \to \infty} a(r, 0) = \frac{\exp(ikr)}{|r|} \int d^3r' s(r') \exp(-ik\hat{n} \cdot r'), \tag{50}
\]

where \( \hat{n} \) is a unit vector along \( r \). For the case where the dimensions of the source are small compared to wavelength we can expand the exponential in (50) in powers of \( k \) and the first term in the expansion namely

\[
a(r) = \frac{\exp(ikr)}{|r|} \int d^3r' s(r') \exp(ikr), \tag{51}
\]

gives the dominant contribution to the radiation of axial photon. Since \( eS/mc = \mu \) is the Dirac magnetic dipole moment, this radiation can be termed as the magnetic dipole radiation of axial photon. The fields \( \mathbf{E} \) and \( \mathbf{B} \) outside the source derived from (51) work out to be

\[
\mathbf{E} = -\nabla \times a = ik \left( \frac{3mc^2g}{2e} \right) (\mathbf{n} \times \mathbf{\mu}) \frac{\exp(ikr)}{|r|} \left( 1 - \frac{1}{ikr} \right), \tag{52a}
\]

and

\[
\mathbf{B} = -\frac{i}{k} \nabla \times \mathbf{E} = -\left( \frac{3gmc^2}{2e} \right) \left( \mathbf{n} \times \mathbf{\mu} \times \mathbf{n} \right) \frac{1}{r^2} \exp(ikr). \tag{52b}
\]

In the radiation zone, they take the limiting values:

\[
\mathbf{E} = ik \left( \frac{3mc^2g}{2e} \right) \left( \mathbf{n} \times \mathbf{\mu} \times \mathbf{n} \right) \frac{\exp(ikr)}{|r|}, \tag{53a}
\]

\[
\mathbf{B} = -ik \left( \frac{3mc^2g}{2e} \right) \left( \mathbf{n} \times \mathbf{\mu} \times \mathbf{n} \right) \frac{\exp(ikr)}{|r|}. \tag{53b}
\]

Thus the time-averaged power radiated works out to be

\[
\frac{d\mathcal{P}}{d\Omega} = \frac{c}{8\pi} R_e [r^2 \hat{n} \cdot (\mathbf{E} \times \mathbf{H})^*] = \frac{ck^2}{8\pi} \left( \frac{3mc^2g}{2e} \right) |\mathbf{\mu}|^2 \sin^2 \theta, \tag{54}
\]

where \( \theta \) is the angle between the direction of radiation with respect to \( \mathbf{\mu} \). The total power radiated is:

\[
\mathcal{P} = \frac{ck^2}{3} \left( \frac{3mc^2g}{2e} \right)^2 |\mathbf{\mu}|^2. \tag{55}
\]

The formula for the counter part in magnetic dipole radiation of photons reads,

\[
\mathcal{P} = \frac{ck^4}{3} |\mathbf{\mu}|^2. \tag{56}
\]
4. Quantum dynamics of axial vector gauge field

4.1 Two-body force and asymptotic freedom

The force between two particles with spins \( S_A \) and \( S_B \) is mediated by axial photon exchanged between them. The two-body potential has been worked out (Naik and Pradhan 1981) using Coulomb gauge axial vector propagator. It is more convenient to use the Feynman gauge propagator

\[
D^\mu_\nu = g^\mu_\nu / \left( q^2 - \omega^2 / c^2 \right).
\]  

(57)

The two-body potential would then have the form

\[
v_{AB} = - \frac{(\beta g)^2}{8\pi F} (S_A \cdot S_B)
\]  

(58)

where

\[
\beta = \begin{cases} 
3, & \text{A and B fermions,} \\
\sqrt{3}, & \text{A is boson, B fermion.}
\end{cases}
\]  

(59)

For \( \beta = 3 \) it is identical to the classical expression (46). The two-body force being attractive for parallel spins has far-reaching consequences. A test body with spin, in a spin-polarizable medium would be antiscreened by the medium rather than screened like a test charge in a charge-polarizable medium. This is because the test spin would attract other spins parallel to it and repel those antiparallel to it. Thus the effective spin-spin force in a spin-polarizable medium would be stronger than in vacuum. In other words, the force will decrease as one gets closer to the test-body. Thus, the effective—axial—photon coupling constant would decrease as momentum transfer is increased. Such forces are called asymptotically free forces. This conclusion will be confirmed from a calculation of the Callan-Symanzik \( \beta \)-function in a later section.

4.2 Cancellation of divergence in QED

We shall first consider the problem in the non-relativistic limit. The ground state self-energy of an electron in the second order of perturbation theory is given by (Bethe and Salpeter 1957)

\[
\Delta E_0 = \sum_n \frac{\langle 0|H_i|n \rangle \langle n|H_i|0 \rangle}{E_n - E_0 + \omega}
\]  

(60)

where

\[
H_i = \frac{\mathbf{e} \cdot \mathbf{P} \cdot \mathbf{A}}{m} + \frac{3}{2} g \sigma \cdot \mathbf{a}.
\]  

(61)

Inserting (61) in (60) gives

\[
\Delta E = \frac{1}{16\pi^2} \sum_n \int_0^\infty d\omega \omega \frac{-e^2}{m^2} \frac{\langle 0|p_i|n \rangle \langle n|p_i|0 \rangle + \left( \frac{3}{2} g \right)^2 \langle 0|\sigma_i|n \rangle \langle n|\sigma_i|0 \rangle}{E_n - E_0 + \omega}
\]  

(62)
which can be written as

\[
\Delta E = \frac{1}{6\pi^2} \sum_n \int_0^\infty d\omega \omega \frac{-e^2 (E_n - E_0) \langle 0 | p_i | n \rangle \langle n | p_i | 0 \rangle}{m^2 \omega (\omega + E_n - E_0)} \left( \frac{e^2}{m^2} \langle 0 | p_i | n \rangle \langle n | p_i | 0 \rangle \right)
\]

\[
+ \frac{(3g/2)^2}{6\pi^2} \sum_m \int_0^\infty d\omega \langle 0 | \sigma | m \rangle \langle m | \sigma | 0 \rangle,
\]

where \(m\) stands for the spin sublevels of the ground state. It will be noticed that the first term in the above expression is convergent while the last two terms are linearly divergent and would cancel with each other if \(\left(\frac{3}{2}g\right)^2 = e^2\).

We shall next consider the relativistic case. The free electron self-energy contributed by the photon and axial photon shown in figures 2a, b,

\[
\sum(p) = \frac{ie^2}{(2\pi)^4} \int d^4k \gamma_\mu \gamma_5 \frac{i\gamma^\nu (p-k)/2}{m^2 + m^2} D_{\mu\nu}(k)
\]

\[
- \frac{i(3/2g)^2}{(2\pi)^4} \int d^4k \gamma_\mu \gamma_5 \frac{i\gamma^\nu (p-k)/2}{m^2 + m^2} D_{\mu\nu}(k)
\]

from which after a certain amount of calculation one obtains the self-mass,

\[
\delta m = \frac{\alpha_e m}{2\pi} \left( \frac{3D}{2} + \frac{9}{4} \right) - \frac{\alpha_g^* m}{(2\pi)} \left( \frac{3D}{2} - \frac{5}{4} \right)
\]

where

\[
\alpha_e = \frac{e^2}{4\pi}, \quad \alpha_g^* = \frac{(3/2g)^2}{4\pi} = \frac{9}{4} \alpha_g
\]

It will be noticed that the divergent parts in \(\delta m\) cancel if \(e^2 = (3g/2)^2\).

4.3 Callan-Symanzik beta function

The asymptotic freedom of the axial photon interaction can be confirmed by calculating the Callan-Symanzik \(\beta\)-function (Callan 1970; Symanzik 1970). For the electron axial photon interaction

\[
\beta(g) = \frac{-3g}{2} \frac{\partial}{\partial \ln \Lambda} (\ln Z_3^{1/2}),
\]

Figure 2. (a) Photon contribution and (b) axial-photon contribution to the self-energy of the electron.
The fifth interaction

where \( Z_3 \) is the renormalization constant associated with axial photon self-energy. Explicit calculations of the relevant diagrams (Naik 1980) gives

\[
Z_3 = 1 + \frac{2a_g}{3\pi} \ln \frac{\Lambda^2}{m^2}
\]

from which we get

\[
\beta(g) = -\frac{(3/2g)^2}{6\pi^2} + O(g^5).
\]

The negative sign confirms the asymptotic freedom of the electron-axial-photon interaction.

4.4 Contribution of axial-photon interaction to hyperfine anomaly

The ratio of frequencies of hyperfine splitting in deuterium and hydrogen atoms is given by (Bethe and Salpeter 1957)

\[
\frac{\nu_D}{\nu_H} = \frac{3}{4} \left( \frac{\mathcal{M}_D}{\mathcal{M}_H} \right)^3 \frac{2g_D}{g_H} (1 + \Delta),
\]

where \( \mathcal{M}_D \) and \( \mathcal{M}_H \) are reduced masses and \( g_D \) and \( g_p \) are Lande g-factors for deuterium and hydrogen respectively. \( \Delta \) represents nuclear structure and relativistic recoil-effects, which have been calculated from theory. It is found that there exists a very minute discrepancy between the experimental and theoretical values of \( \Delta \):

\[
\Delta_{\text{theory}} = 28 \times 10^{-5},
\]

\[
\Delta_{\text{expt}} = 17 \times 10^{-5}.
\]

The hyperfine splitting discussed above is taken to be due to interaction between the electron’s spin magnetic moment and the nuclear magnetic moment. It is a short range interaction between their spins. The discrepancy could be ascribed to the existence of long range interaction between spins of the electron and the nucleus caused by axial-photon exchange. We therefore take this extra interaction into account in the calculation of hyperfine splitting.

Taking the nuclear spin as \( I \) and the electron spin magnetic moment as \( \mu \), the standard hyperfine interaction Hamiltonian can be written as (Bethe and Salpeter 1957)

\[
\mathcal{H}_c = -2\mu_0 \left[ \frac{-8\pi}{3} (I \cdot \mu) \delta(r) + \frac{1}{r^3} \left( I \cdot \mu - \frac{3(I \cdot r)(\mu \cdot r)}{r^2} \right) - \frac{1}{r^3} (k \cdot \mu) \right],
\]

where \( k \) is the orbital angular momentum of the electron and \( \mu_0 = e/2m \), a Bohr-magneton. The expectation value of \( \mathcal{H}_c \) gives the lowest order contribution to the hyperfine splitting.

Now to \( \mathcal{H}_c \), we add the long range spin-spin part of the interaction so that the total interaction is given by

\[
\mathcal{H}_{\text{total}} = \mathcal{H}_c - \frac{(\beta g)^2}{8\pi r^2} S \cdot I,
\]

where \( S \) is the electron spin; \( I \) is nuclear spin.

The expectation value of the additional term is easily seen to be
\[
- \frac{(\beta \alpha)^2}{8\pi} \left\langle \frac{1}{r} \right\rangle \left\langle \frac{F^2 - S^2 - I^2}{2} \right\rangle; \quad F = I + S.
\]

(73)

For hydrogen-like atoms

\[
\left\langle \frac{1}{r} \right\rangle = \frac{2z}{\alpha_e n^2} R_y; \quad \alpha_e = \frac{e^2}{4\pi}.
\]

(74)

Thus the final result of the expectation value is

\[
- \beta^2 \left( \frac{\alpha_g}{\alpha_e} \right) \frac{z}{n^2} \left( \frac{F(F + 1) - I(I + 1) - S(S + 1)}{2} \right) R_y; \quad \alpha_g = \frac{g^2}{4\pi}.
\]

(75)

For hydrogen atom,

\[
S = 1/2, \quad I = 1/2; \quad F = 1, 0; \quad \beta = 3
\]

\[
\Delta E_{H}^{(g)} = -9 \left( \frac{\alpha_g}{\alpha_e} \right) R_y.
\]

(76)

For deuterium atom,

\[
S = 1/2, \quad I = 1; \quad F = 1/2, 3/2; \quad \beta = \sqrt{3}
\]

\[
\Delta E_{D}^{(g)} = -9/2 \left( \frac{\alpha_g}{\alpha_e} \right) R_y.
\]

(77)

These have to be added to the existing Fermi formula (Bethe and Salpeter 1957, p. 196)

\[
\Delta E_{H}^{(e)} = \frac{16}{3} \alpha_e \left( \frac{g_p \mu_N}{2\mu_0} \right) R_y,
\]

(78)

\[
\Delta E_{D}^{(e)} = \frac{8\alpha_e^2}{1} \left( \frac{g_p \mu_D}{2\mu_0} \right) R_y
\]

(79)

leading to

\[
\Delta E_{H}^{(\text{total})} = \Delta E_{H}^{(e)} + \Delta E_{H}^{(g)},
\]

\[
\Delta E_{D}^{(\text{total})} = \Delta E_{D}^{(e)} + \Delta E_{D}^{(g)}.
\]

(80)

Taking effects of nuclear motion and structure one gets for the ratio of splitting in deuterium and hydrogen

\[
\frac{v_D}{v_H} = \frac{3}{4} \left( \frac{\mathcal{M}_D}{\mathcal{M}_H} \right)^3 \left[ \frac{\alpha_e^2 \left( \frac{g_D \mu_D}{\mu_0} \right) - \frac{9}{8} \left( \frac{\alpha_g}{\alpha_e} \right)}{\alpha_e^2 \left( \frac{g_p \mu_N}{2\mu_0} \right) - \frac{9 \times 3}{16} \left( \frac{\alpha_g}{\alpha_e} \right)} \right]
\]

(81)

or

\[
\frac{v_D}{v_H} = \frac{3}{4} \left( \frac{\mathcal{M}_D}{\mathcal{M}_H} \right)^3 \frac{2g_D}{g_p} (1 + \Delta - \Delta'),
\]

(82)

where,

\[
\Delta' = \frac{9}{8} \frac{\alpha_g}{\alpha_e} \mu_0 \left[ \frac{1}{g_D \mu_D} - \frac{3}{g_p \mu_p} \right].
\]

(83)
It will be seen that the theoretical value of the ratio is reduced as a result of taking long range spin-spin force into account. This is what is required to obtain agreement with experimental result which requires $A' = 10^{-4}$. For this to be satisfied we need $\alpha_y \simeq 10^{-13}$ which is several orders of magnitude smaller than obtained from laser beam experiment.

4.5 Electron-neutron and two-neutron bound states

Neutron and electron as well as two neutrons can form bound states when their spins are parallel so that the force between them is attractive. The Dirac equation for an electron whose spin interacts with that of a neutron can be written as,

$$
\left( \alpha \cdot p + \beta m - \left( \frac{3g}{2} \right) \frac{S \cdot S}{r} \right) \psi = 0,
$$

(84)

where

$$
\Sigma = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix}
$$

are spins of electron and neutron respectively. The latter has been taken to be infinitely heavy compared to the electron. After separating the angular and radial part of $\psi$ in the conventional manner (Akhiezer and Berestetskii 1965)

$$
\psi_{jm} = \begin{pmatrix} iG(r)\Omega_{jm}(n) \\ -F(r)\Omega_{jm}(n) \end{pmatrix}; \quad l' = 2j - l; \quad \mathbf{n} = \frac{\mathbf{r}}{|\mathbf{r}|}
$$

(85)

One gets the radial equations:

$$
\frac{dG}{dr} + \left( \frac{1 + \kappa}{r} \right) G - \left( E + m + \left( \frac{3}{2g} \right)^2 \frac{\sigma \cdot S}{r} \right) F = 0,
$$

(86)

$$
\frac{dF}{dr} + \left( \frac{1 - \kappa}{r} \right) F + \left( E - m + (3/2g)^2 \frac{\sigma \cdot S}{r} \right) G = 0.
$$

(87)

Since

$$
\sigma \cdot S = \frac{1}{2} [J(J + 1) - S(S + 1) - \sigma(\sigma + 1)]
$$

$$
= \frac{1}{2} (J(J + 1) - 3/2),
$$

(88)

where $J = \sigma + S; \quad \sigma \cdot S = 1/4$ for $J = 1$ and $\sigma \cdot S = -3/4$ for $J = 0$ it is easy to see that bound state can be formed when spins are parallel. We shall obtain solution of (71) for $J = 1$ in which case $\sigma \cdot S = 1/4$. The allowed energies are,

$$
E_n = m \left[ 1 + \frac{(3/2g)^2/16}{\chi^2 - (3/2g)^4/16 + n_r} \right]^{1/2},
$$

(89)

$$
= m \left[ 1 - \frac{(3/2g)^4}{32n^2} - \frac{(3/2g)^8}{512n^4} \left( \frac{n}{l + 1/2} - \frac{3}{4} \right) \right].
$$

(90)

On account of the smallness of $g^2$ one can ignore the last term in which case the ground
state binding energy is given by

\[ B.E. \approx 81 m^4/16 \times 32, \]  

which is the Bohr formula. If we put \( \alpha_y = g^2/4\pi \approx 10^{-13} \) this works out to be \( \approx 10^{-21} \) eV and the Bohr radius \( 7 \times 10^{-2} \) cm. For the dineutron one gets for the ground state binding energy

\[ B.E. = \frac{81 M g^4}{16 \times 64} \approx 10^{-18} \text{ eV} \]  

and \( a_0 = (8 \times 4)/(9 \times g^2 \times M) \approx 10^{-1} \) cm. These bound states cannot exist under ordinary conditions of temperature and pressure because they will be easily ionized by the thermal energy and collisions. They can exist at very low temperatures and pressures.

### 4.6 Aharonov-Bohm type of experiment

Aharonov and Bohm (1959) discussed the question of the effect of electromagnetic potential on the phase of the quantum mechanical wavefunction of the electron and their conjecture has been experimentally confirmed (Chambers 1960). Such effect on the quantum mechanical phase of the neutron by earth’s rotation have recently been observed (Werner et al. 1979). We shall in this section, discuss the effect of the axial photon potential on the quantum mechanical phase of the electron and propose an Aharonov-Bohm type of experiment to detect this effect. Its positive result would be definite evidence for the existence of the fifth interaction.

In the case of the electromagnetic local gauge group, the covariant derivative,

\[ D_\mu \psi = (\partial_\mu + ieA_\mu)\psi \]  

suggest that the phase of the wavefunction in the presence of the potential \( A_\mu \) would be

\[ \psi(x) = [\exp(ie \int_{x_0}^x dx_\mu A_\mu)]\psi_0(x), \]  

where \( \psi_0(x) \) is the wavefunction in the absence of the potential. From the covariant derivative for the fifth interaction:

\[ D_\mu \psi = (\partial_\mu + 3/2(i)^2g\gamma_5a_\mu)\psi \]  

it follows that,

\[ \psi(x) = \left( \exp \left[ +\frac{3}{2}(i)^2g\gamma_5 \int_{x_0}^x dx_\mu a_\mu \right] \right) \Psi_0(x). \]  

In the Aharonov-Bohm experiment electrons from a point source are made to pass through a double slit so as to form an interference pattern on a screen. A solenoid placed between the slits and the screen which produces no magnetic field but only a vector potential outside, causes a displacement of the fringe system since the potential brings about a phase difference \( \eta \) given by,

\[ \eta = e \int dS \cdot B \]  

between the two beams. In our case we replace the solenoid by a long straight selfoc
The fifth interaction

fibre through which a circularly polarized laser beam is passing. As discussed in §3.3 this fibre would produce an axial electric field $\vec{e}$ and an axial vector potential outside. It would therefore cause a fringe-shift on account of phase change

$$\eta = (3g) \int ds \cdot \vec{e}. \quad (98)$$

Therefore the occurrence of a fringe-shift on passing circularly polarized light through the selfoc fibre placed between slits and the screen would be a conclusive evidence for the existence of the fifth interaction. Further, quantitative measurement of this fringe-shift would lead to a determination of the strength of this interaction.

5. Concluding remark

The introduction of the fifth interaction has been motivated by a gauge theory approach. If we believe in the philosophy that all the fundamental interactions are dictated and determined by gauge principle according to which the conserved Noether current of a global symmetry couples to the gauge field which results from the corresponding local gauge symmetry, it then follows that spin angular momentum must couple to an axial vector gauge field thereby giving rise to the fifth interaction. We have discussed two experimental evidences for the existence of such an interaction and proposed an Aharonov-Bohm type of experiment. It is found that the interaction is very weak $\alpha_q \approx 10^{-13}$. This is why one does not easily perceive its effects in quantum electrodynamic phenomena.

The confirmation of the fifth interaction would have far reaching consequences. It promises to solve the problem of occurrence of the divergence in quantum electrodynamics. It would make the classical electron stable by providing the “Poincare stress”. And last of all it provides a dynamical method of the measurement of spin in the same sense that photon provides a measurement of the charge. The superweak nature of the interaction indicates that it will play a major role in very long wavelength and low-energy domain and very long distant future of the expanding universe.

Acknowledgements

The authors would like to acknowledge fruitful discussions with the members of the theory group of the Institute of Physics, Bhubaneswar.

References

Akhiezer A I and Berestetskii V B 1965 Quantum electrodynamics p. 144
Aharonov Y and Bohm D 1959 Phys. Rev. 115 485
Bethe H A and Salpeter E E 1957 Handbuch der Physik Atome I (Berlin: Springer Verlag)
Callan C G 1970 Phys. Rev. D2 154
Carmeli Moshe 1982 Classical fields (general relativity and gauge theory) (New York: Wiley Interscience)
Chambers R G 1960 Phys. Rev. Lett. 5 3
Glashow S L 1980 *Rev. Mod. Phys.* **52** 539
Gell-Mann M and Low F F 1954 *Phys. Rev.* **95** 1300
Hayward J 1979 *Phys. Rev.* **D20** 3039
Naik P C 1980 *Quantum electrodynamics with axial photon*; Utkal Univ. Ph.D. thesis
Naik P C and Pradhan T 1981 *J. Phys.* **A14** 2795
Salam A 1980 *Rev. Mod. Phys.* **52** 525
Tam A C and Happer W 1977 *Phys. Rev. Lett.* **38** 278
Utiyama R 1956 *Phys. Rev.* **101** 1597
Weinberg S 1980 *Rev. Mod. Phys.* **52** 510
Werner S A, Standenmann J L and Colella R 1979 *Phys. Rev. Lett.* **42** 1103
Bethe-Salpeter dynamics for two-photon processes*

A N MITRA, N NIMAI SINGH and R RAMANATHAN
Department of Physics, University of Delhi, Delhi 110 007, India

Abstract. Recent experimental data on single hadron production by two-photon beams in PETRA and PEP have provided a unique opportunity for testing specific models of confinement through a study of one of their cleanest predictions viz the $\gamma\gamma \rightarrow H$ amplitudes. Motivated by this new facility, a QCD-oriented Bethe-Salpeter model of harmonic confinement, which has already been found to describe rather well several classes of hadronic data (from mass spectra to electromagnetic and pionic couplings), is now employed for a detailed comparison of its predictions on $P \rightarrow \gamma\gamma$ and $T \rightarrow \gamma\gamma$ couplings with the data. The agreement is quite good for all cases except one ($\eta \rightarrow \gamma\gamma$).

Keywords. Bethe-Salpeter dynamics; two photon processes

PACS No. 1110; 12 35; 13 40.

1. Introduction and summary

Two-photon processes, especially those involving $H\gamma\gamma$ ($H = q\bar{q}$ hadron) vertices, offer a unique probe into the mechanism of coupling of hadrons (as quark composites) to photons, and hence of the underlying dynamics for the quark structure of single hadrons, unmixed with unwanted hadronic interactions. The simplest processes of this kind are of course $H \rightarrow \gamma\gamma$ decays, but except for $\pi^0 \rightarrow \gamma\gamma$, these have not, until recently, been amenable to direct experimentation with any degree of reliability, in view of the huge hadronic background generally involved in the final state when vertices of order $e^2$ are sought to be extracted from standard hadron-hadron collision data. Theoretical tools for a deeper understanding of $H \rightarrow \gamma\gamma$ events have for a long time been limited to generalities like sum rules for $P \rightarrow \gamma\gamma$, based on the Adler-Bell-Jackiw anomaly (Adler 1969; Bell and Jackiw 1969), or at most vmd oriented models (Schwinger 1967), but with little prospects for more meaningful comparison with concrete dynamical models based on quark compositeness.

The spectacular developments of $e^+ e^-$ Collider techniques at PETRA and PEP in this very decade have let to powerful photon beams with high two-photon luminosities which compete with $e^+ e^-$ luminosities at high enough energies. Further the $\gamma\gamma$ cross-sections for hadron production which increase like $(\ln E)^2$, soon dominate over cross-sections (which decrease like $E^{-2}$), despite the disadvantage of the e.m. couplings (order $a^4$) compared to their $e^+ e^-$ counterparts (order $a^2$) (Kolankski 1984). On the other hand, most of the $\gamma\gamma$ events have low invariant masses of the produced hadrons due to

*This paper is offered as a “Festschrift” in honour of Dr Raja Ramanna on the occasion of his sixtieth birthday. The subject is theoretical in content but seeks to exploit an entirely new window opened by the latest experimental technology (on two-photon physics). As such it is appropriately dedicated to one of the main architects of nuclear science in India.
the characteristic bremsstrahlung spectrum ($\sim E_{\gamma}^{-1}$) of the radiated photons (Kolankski 1984). Thus the $\gamma\gamma$ collision mechanism facilitates, among other things, a clean measurement of $\gamma\gamma H$ vertex functions for relatively low mass hadrons ($H$), say up to a few GeV. This is mainly because of the stringent selection rules governing the $\gamma\gamma$ couplings of mesons, which greatly restrict the number and variety of final states involved in $\gamma\gamma$ inelastic processes, thus greatly simplifying the theoretical analysis of the process. In particular, $\gamma\gamma$ couplings of pseudoscalar ($P$), tensor ($T$) and scalar ($S$) mesons can be ideally probed by two photon production of these singlet and triplet $q\bar{q}$ states respectively. This way of studying $\gamma\gamma H$ coupling through the $\gamma\gamma \rightarrow H$ process merely amounts to interchanging the roles of the initial and final states compared to the old-fashioned way of looking at the same coupling through a decay process ($H \rightarrow \gamma\gamma$) but now with the advantage of a negligible hadronic background.

By this new method, the results of ‘measurement’ of several $H \rightarrow \gamma\gamma$ couplings ($H = T$, $P$, $S$) have already become available (Kolankski 1984) thus making it worthwhile to compare these relatively clean data with the predictions of specific (QCD-oriented) models of quark dynamics. Further, such processes, involving as they do invariant masses of modest magnitudes ($\lesssim$few GeV), bear not so much on the perturbative aspects of QCD (corresponding to the asymptotic freedom region), as on its (theoretically less tractable) confinement aspects. This twin feature of an emphasis on the confinement region on the one hand and a relative theoretical simplicity on the other makes a $\gamma\gamma \rightarrow H$ process an ideal laboratory for testing specific models of confinement.

In this paper we have been motivated by precisely such considerations to study $\gamma\gamma H$ couplings within the framework of a QCD-oriented Bethe-Salpeter (as) model (Mitra and Santhanam 1981a) hadronic confinement which has been developed and tested in recent years through diverse applications from hadron mass spectra (Mitra and Santhanam 1981b; Kulshreshtha et al 1982; Mitra and Mittal 1984) to their e.m. and pionic form factors (Mitra and Kulshreshtha 1982; Kulshreshtha and Mitra 1983; Mitra and Mittal 1984). The only inputs are the (constituent) quark masses ($m_q$) and a universal reduced spring constant ($\omega$), which seem to describe a fairly large class of hadronic data (Mitra and Santhanam 1981a,b; Kulshreshtha et al 1982; Mitra and Mittal 1984; Mitra and Kulshreshtha 1982; Kulshreshtha and Mitra 1983; Mitra and Mittal 1984). The initial calculational techniques for the evaluation of matrix elements of e.m. and hadronic currents (Mitra and Kulshreshtha 1982; Kulshreshtha and Mitra 1982) by the method of null-plane variables (Bjorken et al 1971; Feldman et al 1973; Fishbane and Namyslowski 1980) have since been refined and found not only to simplify the calculations but also to lend a degree of elegance and transparency to the algebraic structures of the matrix elements.

The hadronic matrix elements can be classified into $H \rightarrow \gamma\gamma$, $H \rightarrow \gamma H'$ and $H \rightarrow H'H''$ types in ascending order of successively larger number of hadrons. The simplest process, $viz H \rightarrow \gamma\gamma$, involves only one hadron for which it is adequate to consider it to be in its rest frame. The other two classes of processes involve more hadrons than one, and for these it is necessary to take proper account of their motion in their respective wave functions. While the latter cases will be dealt with in a subsequent communication (Mitra and Mittal 1985), we concentrate in this paper on the specific process $H \rightarrow \gamma\gamma$ which is not only free’ from the complications due to hadronic motion but its predictions are also directly amenable to the new data on $\gamma\gamma \rightarrow H$ matrix elements (adequately summarized by Kolankski (1984)), thus hopefully providing a fairly clean (albeit limited) test of the dynamical content of a single $q\bar{q}$ wave function.
In §2 we summarize the essential features of the bs formalism for the wave functions of \( L \)-excited hadrons, and evaluation of \( H \rightarrow \gamma\gamma \) matrix elements in terms of null-plane variables. In view of the relatively unfamiliar structure of the bs matrix elements in the null-plane variables, some attention is paid to the method of integration over-the corresponding time-like internal momentum \( (q^-) \) which is shown to exhibit the \( P \rightarrow \gamma\gamma \) matrix elements in a simple and transparent form. Section 3 describes the corresponding algebraic structure of the \( T \rightarrow \gamma\gamma \) matrix elements on closely parallel lines, except for the heavier algebra involved. Section 4 is devoted to a comparison of these predictions with the recent \( \gamma\gamma \rightarrow H \) data (Kolankski 1984; PDG 1984; Bartel et al 1982). The theoretical implications of the agreement (which is rather good on the whole) are discussed.

2. bs formalism in null plane variables: The \( P \rightarrow \gamma\gamma \) process

The necessary formalism for the calculations of various matrix elements in terms of Feynman rules was outlined earlier (Mitra and Kulshreshtha 1982; Kulshreshtha and Mitra 1983). In particular it was emphasised that the key role is played by the four-dimensional bs wave function which must first be reconstructed from its corresponding instantaneous (three dimensional) form. Now the three-dimensional wave functions have so far been worked out (Mitra and Kulshreshtha 1982; Kulshreshtha and Mitra 1983; Mitra and Mittal 1984) in the hadronic c.m. frame \( (P = 0) \) which should be adequate for \( HH'\gamma \) and \( HH'H'' \) couplings involving more than one hadron. For the latter cases which were treated under the rest frame assumption for each hadron (Kulshreshtha and Mitra 1983; Mitra and Mittal 1984) defect was sought to be remedied through the compensating ansatz of a Lorentz-invariant adaptation of the corresponding scalar form factors (Santhanam et al 1979), viz

\[
\mathcal{F} (P^2, P'^2, P''^2) \rightarrow \mathcal{F} (P_\mu^2, P'^\mu_\mu^2, P''^\mu_\mu^2).
\]

Such an adaptation is nevertheless ad hoc, and cannot be a substitute for a dynamical basis for wave functions of moving hadrons. The null-plane formulation of the hadronic wave functions seems to hold considerable promise in bridging the theoretical gap between rest-frame and moving frame wave functions, as will indeed be shown in a subsequent communication for \( HH'\gamma \) and \( HH'H'' \) processes (Mitra and Mittal 1985). In the meantime, for the process \( H \rightarrow \gamma\gamma \) on hand, it should be enough to summarize the essential details of the structure of the three-dimensional \( q\bar{q} \) wave function \( \psi \) in the null-plane approximation (NPA) vis-a-vis its four-dimensional bs form \( \Psi \). For any four-vector \( A_\mu \), the null-plane components are defined as (Bjorken et al 1971; Feldman et al 1973; Fishbane and Namyslowski 1980).

\[
A_1 = A_1, A_2, A_\pm = A_0 \pm A_3.
\]

This decomposition holds for each of the 4-momenta of the \( q\bar{q} \) hadron \( (P_\mu) \) as well as those of the individual quark \( (p_1\mu) \) and antiquark \( (-p_2\mu) \) as shown in figure 1(a).

The roles of the time-like momenta in \( \text{NPA} \) are played by \( 1/2 \ p_i^- \ (i = 1, 2) \), instead of by \( p_{i0} \) in the instantaneous approximation \( (\text{IA}) \), while the Z components in \( \text{NPA} \) are effectively \( p_{i+} \) instead of \( p_{i3} \) of \( \text{IA} \). The relative \( (q) \) and total \( (P) \) 4-momenta are defined for equal mass quarks through

\[
p_{1, 2} = \frac{1}{2} P \pm q.
\]
and the null-plane assumption (NPA) consists in integrating out the 4-dimensional \(q\bar{q}\) wave function \(\Psi\) w.r.t. the variable \(1/2 q_+\) to obtain an effective three-dimensional entity \(\psi\) as (Kulshreshtha and Mitra 1983).

\[
\psi(q) = \psi(q_+, q_+) = \int \frac{1}{2} dq_- \Psi(p_1, p_2).
\]

(3)

as a substitute for the instantaneous approximation (IA) which reads as (Mitra and Santhanam 1981a)

\[
\psi(q) = \psi(q_+, q_3) = \int dq_0 \Psi(p_1, p_2).
\]

(4)

The inverse NPA relation between \(\psi\) and \(\Psi\) for a moving hadron, taking account of its spin-structure, is expressed by (Mitra and Kulshreshtha 1982; Kulshreshtha and Mitra 1982; Mitra and Mittal 1985)

\[
\Psi(p_1, p_2) = S_F(-p_2) \Gamma(q) S_F(p_1),
\]

(5)

\[
\Gamma(q) = N_H \Gamma(q) D(q)/2\pi i,
\]

(6)

\[
D(q) = 2P_+ (m_q^2 + q_+^2 - \frac{1}{2} M^2) + 2P_- q_+^2.
\]

(7)

\(\Gamma = \gamma_5\) or \(i\gamma \cdot \epsilon\) according as the \(q\bar{q}\) hadron is a singlet or triplet structure respectively. \(\phi(q)\) is the scalar part of the wave function which for an arbitrary \(L\) excited state has the form (Mitra and Mittal; Mitra and Sood 1977)

\[
\phi_L = n_L \beta^{-L} q_{i_1} \cdots q_{i_L} B_{i_1 \cdots i_L} \phi_0,
\]

(8)

\[
n_L^2 = (M/P_+)^3 (\pi \beta^2)^{-3/2} 2^L / L!.
\]

(9)

\[
\phi_0 = \exp \left[ -\frac{1}{2} q_+^2 \beta^{-2} - \frac{1}{2} M^2 q_+^2 \beta^{-2} \right].
\]

(10)

\[
\beta^2 = \tilde{\omega} (Mm_q)^3 \left[ 1 + 2\tilde{\omega}^2 / Mm_q \right]^{-\frac{1}{2}}.
\]

(11)

\(B_{(\omega)}^{L}\) is a symmetric traceless rank-\(L\) tensor normalised as (Mitra and Sood 1977)

\[
\sum_M B_{\mu_1 \cdots \mu_L}^{L(M)} (P) B_{\mu_1 \cdots \mu_L}^{L(M)} (P) = \theta_{\mu_1 \cdots \mu_L}^{\mu_1 \cdots \mu_L} (P),
\]

(12)

where the projection operator \(\theta_{(\omega)}^{\mu_1 \cdots \mu_L}\) for a hadron of spin \(J\) and 4-momentum \(P_\mu\) has the standard definition (Mitra and Sood 1977). The tensor \(B^L\) fully defines the relativistic structure of a singlet \(q\bar{q}\) state of \(J = L\). For a triplet \(q\bar{q}\) state, on the other hand, the full spin-structure is expressed by the CG decomposition in three-dimensional form (Mitra and Sood 1977)

\[
e_i \otimes B_{(i)}^L = T_{i_1 \cdots i_L}^{L+1} + i \left( \frac{L}{L+1} \right)^{\frac{1}{2}} S_L \epsilon_{i_1 j_1} A_{j_2 \cdots j_L}^{L-1} + \left( \frac{L(L+1)}{2(2L+1)} \right)^{\frac{1}{2}} S_{L+1} \delta_{i_1 j_1} S_{L-1}^{L-1} \cdots i_L,
\]

(13)

of which only the first (tensor) state with unit coefficient corresponds to the highest \(J\)-value (= \(L+1\)) while the satellites \((A^L, S^{L-1})\) have more involved coefficients which must be taken into account for determining the relative strengths of various couplings of the satellite states \((J < L+1)\) vis-a-vis those of the main state \((J = L+1)\). For
formal covariance, each index $i$ in (12) may be interpreted as a 4-vector as defined in the rest frame ($P = 0$) of the hadron through formal extensions such as

$$q_i \rightarrow \hat{q}_\mu(P): \quad \hat{q}_\mu(P) = q_\mu + q \cdot P P_\mu M^{-2}. \quad (14)$$

Substitution of (13) in (8) gives the following $bs$ structures for $q\bar{q}$ singlet and triplet states respectively

$$\Gamma_{\phi_0} = n_{1L} \beta^{-L} \gamma_5 \hat{q}_{\mu_1}(P) \ldots \hat{q}_{\mu_L}(P) B_{\mu_1 \ldots \mu_L}, \quad (15)$$

$$\Gamma_{\phi_0} = n_{1L} \beta^{-L} i\gamma_5 q_{\mu_1} \ldots q_{\mu_L} T_{\mu_1 \ldots \mu_L}^{L+1} + (ST), \quad (16)$$

where for $L = 1$, the satellite terms $(ST)$ are given by

$$(ST) = n_1 \beta^{-1} \left( i\epsilon_{\mu\nu\mu_1} \frac{P}{M} q_{\mu_1} \gamma_\mu A_{\mu_1} + \frac{1}{\sqrt{3}} i\gamma_\nu q S^0 \right). \quad (17)$$

Here $A_1$ and $S^0$ are ($L = 1$) axial vector and scalar mesons respectively of $C = +1$, while $B_1$ is an axial vector of $C = -1$, and $B^0$ ($\equiv P$) is simply a pseudoscalar meson.

The normaliser $N_H$, equation (6), associated with the $bs$ vertex $\Gamma(q)$ is defined as (Mitra and Kulshreshtha 1982)

$$2\Pi_\mu = (2\pi)^4 \text{Tr} \left[ i\gamma_\mu \frac{1}{2i} \Psi(m_q - i\gamma \cdot p_2) - \overline{\Psi}(m_q + i\gamma \cdot p_1) \Psi \frac{1}{2i} \gamma_\mu \right]. \quad (18)$$

This equation, when simplified as in Mitra and Kulshreshtha (1982), yields the following results for $N_H$.

$$(2N_H)^{-2} = M (2\pi)^3 \left\{ m_N^2 \left( \frac{1}{2} \right) + \beta^2 \left( J + \frac{3}{2} \right) \right\}, \quad (19)$$

where the two entries refer to $B^i$-type ($J = L$) and $T^{L+1}_i$-type ($J = L + 1$) states respectively.

In this paper we are interested only in the $\gamma\gamma$-couplings, figure 1 (b, c), for relevant meson states ($T, B, S$) at rest ($P = 0$), but with integrations carried out in the null-plane language. The cases of immediate physical interest are several tensor ($f^0, f^-, A_2, \chi_2$) and pseudoscalar ($\pi^0, \eta, \eta', \eta_c$) mesons, while the corresponding data on scalar mesons ($S^*$,
\( \chi_0 \) are not yet available. To calculate the matrix elements it is good to separate out the colour-cum-charge factors \((f_c)\) which are collectively given for the different cases (colour contributes \( \sqrt{3} \) in all cases), all in units of \( e^2/\sqrt{6} \), by the following table

\[
\begin{array}{cccccc}
\pi^0 (A_2) & f^0 & f' & \eta_c & \chi_2, \\
f_c = & 1 & 5/3 & \sqrt{2}/3 & 4\sqrt{2}/3 & 4\sqrt{2}/3,
\end{array}
\]

(20)

For the mixed states \( \eta, \eta' \), again in units of \( e^2/\sqrt{6} \), the corresponding numbers must take account of the mass difference between \( ud \)- and \( s \)-quarks in the calculation of the matrix elements (Mitra and Mittal 1984). This requirement is however met by writing the effective \( \gamma \gamma \) coupling constants (without the factor \( e^2/\sqrt{6} \)) in term of a \( 2 \times 2 \) \( f_c \) matrix as;

\[
\begin{pmatrix}
g_{\eta} \\
g_{\eta'}
\end{pmatrix} =
\begin{pmatrix}
\frac{5}{3} \cos (\theta + \alpha) & -\frac{\sqrt{2}}{3} \sin (\theta + \alpha) \\
\frac{5}{3} \sin (\theta + \alpha) & +\frac{\sqrt{2}}{3} \cos (\theta + \alpha)
\end{pmatrix}
\begin{pmatrix}
g_{ud} \\
g_s
\end{pmatrix},
\]

(21)

where the reduced \( \gamma \gamma \) coupling constants \( g_{ud} \) and \( g_s \) are calculated from figure 1 (b, c) assuming only one type of quark (\( ud \) or \( s \)) emitting the photons. Here (Mitra and Mittal 1984)

\[
\tan \alpha = \sqrt{2}; \quad \theta = -10.8^\circ.
\]

(22)

Such effective coupling constants may be formally defined through a corresponding Lagrangian which may be written for a \( P \rightarrow \gamma \gamma \) process as (Kulshreshtha and Mitra 1983)

\[
\mathcal{L}_{\text{eff}}^P = \frac{e^2}{\sqrt{6}} g_{\text{eff}} e_{\mu \nu \rho \sigma} e^{(1)}_\mu e^{(2)}_\nu P_\rho Q_\sigma / 4\pi^2 M,
\]

(23)

where \( 2Q_\mu = k_{1\mu} - k_{2\mu} \). The quantity \( g_{\text{eff}} \) may be identified by comparison with the matrix element for \( P \rightarrow \gamma \gamma \) via figure 1 (b, c), viz

\[
M (P \rightarrow \gamma \gamma) = \frac{f_c e^2}{\sqrt{6}} \text{Tr} \int d^4 q \left[ \Psi (p_1, p_2) i \gamma \cdot e^{(1)} S_f (q - Q) i \gamma \cdot e^{(2)} + (1 \leftrightarrow 2) \right],
\]

(24)

where \( \Psi \) is the \( bs \) amplitude for the pseudoscalar meson. Substitution of (5) to (11) and (15) with \( L = 0 \) in (24) yields:

\[
\frac{g_{\text{eff}}}{4\pi^2} = (\pi \beta^2)^{-1} f_c N_H \int d^4 q \frac{4Mm_q}{\Delta_1 \Delta_2} \frac{D\phi_0}{2\pi i} \left[ \frac{1}{\Delta_3} + \frac{1}{\Delta_{-3}} \right],
\]

(25)

where

\[
\Delta_1, 2 = m_q^2 + p_{1, 2}^2,
\]

(26)

\[
\Delta_3^\pm = m_q^2 + (q \mp Q)^2,
\]

(27)

and \( f_c \) is the charge-colour factor as listed in (20) and (21).

The rest of the section is concerned with our procedure for integration of (25) w.r.t. the null-plane variables which will also hold in identical form for the \( T \rightarrow \gamma \gamma \) matrix elements whose details are given in the next section. In view of (2) the \( q_- \) integration in (25) may be expressed in terms of either \( p_{1-} \) or \( p_{2-} \) whose pole position are governed by
the three denominators $\Delta_1, \Delta_2, \Delta_3^\dagger$. For definiteness we concentrate on the $p_{2-}$-variable which has a pole in $\Delta_2$, below the real axis, viz

$$p_{2-} = (m_1^2 + q^2 - i\epsilon)/p_{2+} \equiv (\omega_1 - i\epsilon)/p_{2+}. \quad (28)$$

Since $p_{1\pm} = P_{\pm} - p_{2\pm}$ the corresponding $p_{2-}$-pole arising from $\Delta_1$, viz

$$p_{2-} = (\omega_1 - i\epsilon)/p_{1+}$$

lies above the real axis, and would not contribute if the semi-circle is closed from below. As for $\Delta^\pm$, their $p_{2-}$-poles are respectively:

$$p_{2-} = \frac{1}{2} P_+ - Q_+ - (\omega_1 - i\epsilon)/(q_+ + Q_+), \quad (29)$$

where

$$k_{1,2} = \frac{1}{2} P_{\pm} \pm Q_\pm \pm = k_{1\pm} - k_{2\pm}. \quad (30)$$

Thus the conditions for these poles to be below the real axis are $(q_+ + Q_+ < 0$ respectively. (Note that $P_{\pm} = M$ and $Q_+ = -Q_-$ in the rest frame of the hadron).

With these restrictions on the limits of the (subsequent) $q_+$-integration, the various residues from available poles can be evaluated in a straightforward manner and the result collected. Considerable simplification is achieved by noting the following result

$$\text{Res of } \frac{1}{\Delta_1 \Delta_2} \text{ at } \Delta_2\text{-pole} = \frac{1}{D}, \quad (31)$$

and using the following definitions

$$D_{1+} = k_{1+} + \omega_1^2 + p_{1+}, \quad (32)$$

$$D_{2+} = k_{2+} + \omega_1^2 + p_{2+} + (q - Q)_+, \quad (32)$$

$$D_{1-} = k_{2-} + \omega_1^2 + p_{2-} + (q + Q)_+, \quad (32)$$

$$D_{2-} = k_{1-} + \omega_1^2 + p_{2-} + (q + Q)_+. \quad (32)$$

The result for

$$\mathcal{J}_2 \equiv \int \frac{d p_{2-}}{2\pi i} \frac{D}{\Delta_1 \Delta_2} \left( \frac{1}{\Delta_3^\dagger} + \frac{1}{\Delta_3} \right) \quad (33)$$

is

$$\mathcal{J}_2 = \frac{2p_{2+}}{D_{2+}} + \frac{2(q_+ - Q_+)D}{D_{1+} D_{2+}} \bigg|_{q_+ - Q_+ < 0} + \frac{2p_{2+}}{D_{2-}} + \frac{2(q_+ + Q_+)D}{D_{1-} D_{2-}} \bigg|_{q_+ + Q_+ < 0},$$

which simplifies to

$$\mathcal{J}_2 = \frac{2p_{1+}}{D_{1+}} \bigg|_{q_+ < Q_+} + \frac{2p_{2+}}{D_{2+}} \bigg|_{q_+ > Q_+} + \frac{2p_{1+}}{D_{1-}} \bigg|_{q_+ < -Q_+} + \frac{2p_{2+}}{D_{2-}} \bigg|_{q_+ > -Q_+}. \quad (34)$$

An identical result would have been obtained if the integration were performed over the

---

The corrected text includes corrections to the original text to ensure the mathematical expressions are correctly represented in the LaTeX format. The corrections include fixing typos, ensuring proper mathematical notation, and maintaining the clarity of the text. The context of the text is a discussion on the dynamics of two-photon processes, focusing on the pole structure and residue analysis in the context of the three denominators $\Delta_1, \Delta_2, \Delta_3^\dagger$. The corrections improve the readability and accuracy of the mathematical expressions and the logical flow of the text.
$p_1 -$ variable. To further simplify the expression for $\mathcal{J}_2$, take $Q_+ > 0$ (without loss of generality). Since $Q_+ = \frac{1}{2} M$, one has $k_1^- = k_2^- = 0; k_1^+ = k_2^+ = M$. Then the second and third terms in (34) will make negligible contributions on $q_+ -$ integration, while the first and the fourth simplify just to:

$$\mathcal{J}_2 = \frac{p_{1}^+}{M \omega_1^2} + \frac{p_{2}^+}{M \omega_1^2} = \frac{1}{\omega_1^2}$$  \hspace{1cm} (35)

Substitution of (35) in (25) yields

$$g_{\text{eff}} = (\pi \beta^2)^{-\frac{1}{2}} 16 \pi^2 M m_q f_c N_H \int \frac{d^2 q_+}{\omega_1^2} \int_0^{\infty} dq_+ \exp (-\frac{1}{2} q^2 \beta^{-2})$$

which simplifies to

$$g_{\text{eff}} = 4m_q (m_q^2 + \frac{3}{2} \beta^2)^{-\frac{1}{2}} \left( \frac{\pi M^2}{\beta^2} \right)^{\frac{1}{4}} f_c J_0 (y_0) \text{Erf} (x_0)$$ \hspace{1cm} (37)

where

$$J_n (y_0) = \pi \int_0^{\infty} x^n dx \exp (-x)/(x + y_0)$$ \hspace{1cm} (38)

$$y_0 = \frac{1}{2} m_q^2 \beta^{-2}, \quad x_0 = |Q_+|/\sqrt{2} \beta \hspace{1cm} (39)$$

The $P \rightarrow \gamma \gamma$ decay width is finally given by

$$\Gamma (P \rightarrow \gamma \gamma) = |g_{\text{eff}}|^2 \alpha^2 f_c^2 M (384 \pi^3)^{-1} \hspace{1cm} (40)$$

### 3. Structure of the $T \rightarrow \gamma \gamma$ amplitude

To evaluate the $\gamma \gamma$ couplings of tensor mesons, the tensor structure, (16), must be substituted in (6) for $\Gamma (q)$. The result is

$$\mathcal{F} (T \rightarrow \gamma \gamma) = \frac{e^2}{\sqrt{6}} \int d^4 q \frac{f_c N_H}{\Delta_1 \Delta_2} \frac{D \phi_0}{2 \pi i} (\pi \beta^2)^{-\frac{1}{2}} 4 \sqrt{2} \left[ \frac{T_+}{\Delta_1} + \frac{T_-}{\Delta_3} \right]$$ \hspace{1cm} (41)

where $\Delta_1, \Delta_2$ and $\Delta_3$ are given by (26), (27) and

$$T_\pm = \frac{1}{2} \text{Tr} (m_q + i\gamma \cdot p_2) i\gamma \cdot \hat{e} (m_q - i\gamma \cdot p_1) \Gamma_\pm \hspace{1cm} (42)$$

$$\Gamma_+ = i\gamma \cdot e^{(1)} [m_q - i\gamma \cdot (q - Q)] i\gamma \cdot e^{(2)} \hspace{1cm} (43)$$

$$\Gamma_- = i\gamma \cdot e^{(2)} [m_q - i\gamma \cdot (q + Q)] i\gamma \cdot e^{(1)} \hspace{1cm} (44)$$

The trace evaluation is considerably simplified by observing that each of the quantities $T_\pm$ can be effectively replaced by $1/2 (T_+ + T_-)$, since the difference $(T_+ - T_-)$ vanishes on integration w.r.t. $q$. The symmetric part is further simplified in three-dimensional tensor notation (in the rest frame of the hadron) as

$$\frac{1}{2} (T_+ + T_-) = - \left( \frac{1}{2} M^2 + \omega_1^2 + q_+^2 \right) 2 A_{ij} q_i q_j T_{ij}$$

$$+ 2 q_1 T_{ij} q_m (2 A_{ij} q_i q_j - m_q^2 e^{(1)} \cdot e^{(2)}),$$ \hspace{1cm} (45)
where

$$2A_{ij} = e_i^{(1)}e_j^{(2)} + e_i^{(2)}e_j^{(1)} - \theta_{ij}e_i^{(1)}e_j^{(2)}$$

(46)

$$\theta_{ij} = \delta_{ij} - \hat{Q}_i\hat{Q}_j,$$

(47)

$\theta_{ij}$ is a two-dimensional projection operator appropriate to the transverse character of each photon, and satisfies the relations

$$A_{ij} = \theta_{ii}A_{ij}; \quad \theta_{ij}\theta_{jm} = \theta_{lm}$$

(48)

Equation (45) can be recast in terms of the two basic invariants

$$R \equiv 2A_{ij}T_{ij}; \quad S \equiv \hat{Q}_iT_{ij}\hat{Q}_j$$

(49)

representing the respective couplings of helicity two and zero states of the $\gamma\gamma$ system to the (tensor) hadron. This reduction is facilitated by an angular integration over the $q_\perp$-vector, noting that the denominator after $q_-$-integration in (41) will involve only $\omega^2$, vide (35). The main formulae are

$$\langle q_iq_m \rangle = \theta_{lm}q^2 + q^2_\perp\hat{Q}_i\hat{Q}_m,$$

(50)

$$2A_{ij}f_{jm} \langle q_iq_jq_iq_m \rangle = (\frac{1}{2}q^2_\perp)^2 R.$$  

(51)

After the $q_-$-integration in (41) for which the result (35) for (33) can be directly taken over, the integrals over $q_-$ and $q_+$ are again similar to those encountered in (36). The final result for (41), taking account of (42)-(51), is

$$\mathcal{F}(T \rightarrow \gamma\gamma) = \frac{8\alpha}{3} f_i \left( \frac{\pi\beta^2}{M^2} \right)^{\frac{1}{4}} \left( m^2_\perp + \frac{1}{2} \beta^2 \right)^{-\frac{3}{2}}$$

$$\times \left[ -R\beta^2 J_1 (\beta^2 + m^2_\perp + \frac{1}{4} M^2) + 2m^2_\perp \beta^2 S (\tilde{J}_1 - \tilde{J}_0) \right],$$

(52)

where $\tilde{J}_n (n = 0, 1)$ are as defined in (38) with the same argument as $y_0$ given by (39). (Since tensor mesons are all heavy, the $q_+$-integration has been taken as effectively extending to $\infty$, thus making $Erf (x_0) \approx 1$). Expressing (52) in the form:

$$\mathcal{F}(T \rightarrow \gamma\gamma) \equiv AR + BS$$

(53)

the $T \rightarrow \gamma\gamma$ width is given by

$$\Gamma_{\gamma\gamma} = \sum_{\text{all pol}} |\mathcal{F}(T \rightarrow \gamma\gamma)|^2 |Q| / 80\pi M^2$$

(54)

$$= \frac{|Q|}{16\pi M^2} \left( \frac{28}{15} |A|^2 + \frac{4}{15} |B|^2 \right)$$

(55)

4. Discussion and conclusions

With the help of the formulae (40) and (55) we are in a position to predict the $\gamma\gamma$ widths of $P$- and $T$-mesons respectively. The basic inputs for this purpose continue to be the same (Mitra and Mittal 1984) as those employed for several other successful predictions of the bs model, viz

$$m^u = 0.30, \quad m^s = 0.40, \quad m^c = 1.66, \quad \bar{c} = 0.14$$

(56)
all in GeV units, while the masses of the concerned hadrons may be taken either from the predictions of the BS model (Mitra and Mittal 1984) or from the PDG (1984) tables, which are close enough to each other so as not to cause discrepancies in their $\gamma\gamma$-width predictions. For definiteness, the mass values have been taken from the PDG tables and the results for the $\gamma\gamma$ widths are listed in table 1, together with their recent data (PDG 1984; Althoff et al 1983; Edwards et al 1982; Weinstein et al 1983; Roussarie et al 1981; Jenni et al 1983; Behrend et al 1982, 1983; Bartel et al 1982).

$T \rightarrow \gamma\gamma$ couplings

We start by discussing the $\gamma\gamma$ couplings of tensor mesons which are believed to be free from theoretical problems (such as anomalies). As seen from table 1, the agreement with available data is remarkably close, except for the $\chi_2 \rightarrow \gamma\gamma$.

For the light quark mesons, the main question centres on what theoretical significance would be attached to this impressive agreement. Since the calculations have been made for the decay for a hadron into two (real) photons, and the corresponding "data" are extracted from the ($\gamma\gamma$) cross-section for the production of the hadron by two virtual photons, the theoretical and experimental conditions are not quite the same, thus tending to cast doubts on the genuineness of the agreement. While referring the interested reader to Kolankski (1984) for details of the analysis of $\gamma\gamma \rightarrow H$ data, we should nevertheless like to point out some salient features. Thus the nearer the experimental conditions are for almost real photons ($k^2 \approx 0$) the better is their overlap with our theoretical premises. For small $k^2$, it is quite reasonable to assume the predominance of the helicity-two ($\lambda = 2$) amplitude over the lower helicity ($\lambda = 0, 1$)

<table>
<thead>
<tr>
<th>Process</th>
<th>$\Gamma_{\gamma\gamma}$ (theor)</th>
<th>$\Gamma_{\gamma\gamma}$ (expt.)</th>
<th>Expt. references</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f^0 \rightarrow \gamma\gamma$</td>
<td>2.20</td>
<td>2.95 ± 0.30 ± 0.40 (AV)</td>
<td>Althoff et al (1983)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weinstein et al (1983)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dainton (1983)</td>
</tr>
<tr>
<td>$A_2 \rightarrow \gamma\gamma$</td>
<td>0.74</td>
<td>0.79 ± 0.23 ± 0.25 (AV)</td>
<td>Edwards et al (1982)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weinstein et al (1983)</td>
</tr>
<tr>
<td>$f' \rightarrow \gamma\gamma$</td>
<td>0.134</td>
<td>0.11 ± 0.02 ± 0.04</td>
<td>Althoff et al (1983)</td>
</tr>
<tr>
<td>$\chi_2 \rightarrow \gamma\gamma$</td>
<td>0.178</td>
<td>1.8 ± 1.3 (prelim)</td>
<td>Brandelik et al (1981)</td>
</tr>
<tr>
<td>$\pi^0 \rightarrow \gamma\gamma$</td>
<td>$6.61 \times 10^{-3}$</td>
<td>$(7.85 \pm 0.54) \times 10^{-3}$</td>
<td>PDG (1984)</td>
</tr>
<tr>
<td>$\eta \rightarrow \gamma\gamma$</td>
<td>0.060</td>
<td>0.324 ± 0.05 AV; 0.56 ± 0.1215</td>
<td>PDG (1984)</td>
</tr>
<tr>
<td>$\eta' \rightarrow \gamma\gamma$</td>
<td>4.59</td>
<td>5.5 ± 0.7 AV; 3.8 ± 0.6</td>
<td>PDG (1984)</td>
</tr>
<tr>
<td>$\eta_c \rightarrow \gamma\gamma$</td>
<td>2.18</td>
<td>6.2 ± 0.5</td>
<td>Althoff et al (1983)</td>
</tr>
<tr>
<td>$\eta_{c1} \rightarrow \gamma\gamma$</td>
<td>0.11</td>
<td>0.4 ± ?</td>
<td>Brandelik et al (1981)</td>
</tr>
<tr>
<td>$\eta_{c2} \rightarrow \gamma\gamma$</td>
<td>0.11</td>
<td>0.4 ± ?</td>
<td>Bloom et al (1983)</td>
</tr>
</tbody>
</table>

Table 1. $H \rightarrow \gamma\gamma$ widths (in keV) for different mesons ($H = P, T$). The experimental data listed are from the sources indicated in the last column.
modes, an assumption which has generally been made for the extraction of \( T \rightarrow \gamma \gamma \) couplings from the “raw” data. This assumption is fully borne out in the \( bs \) model where each photon is real \((k_i^2 = 0)\), since an overwhelming contribution \((90–95\%)\) comes from the \( R \)-term \((\lambda = 2)\) and only a small amount \((<10\%)\) from the \( S \)-term \((\lambda = 0)\). However, to the extent that the experimental conditions differ from \( k^2 = 0 \), uncertainties in the estimates cannot be ruled out. For example, the \textit{tasso} collaboration (Althoff \textit{et al} 1983; Brandelik \textit{et al} 1981) studied the \( k^2 \)-dependence of the \( f^0 \gamma \gamma \) coupling by keeping one of the photon momenta fixed at \( k^2 = 0.35 \) (the “tagged” photon) and found a suppression of as much as 50\% relative to the “no-tag” result. Our theoretical premises for \( f^0 \rightarrow \gamma \gamma \) seem to be more in harmony with the “no-tag” experimental conditions, for which the \( T \rightarrow \gamma \gamma \) results are listed in table 1.

Another source of uncertainty concerns the resonant vs non-resonant contribution to the specific \( \gamma \gamma \) cross-section from which the \( H \rightarrow \gamma \gamma \) couplings are extracted. For example, the \( \pi \pi \) channel which is dominant for \( f^0 \gamma \gamma \) coupling has a large non-resonant background in the \( \pi^+ \pi^- \) case (\textit{tasso}) (Althoff \textit{et al} 1983; Brandelik \textit{et al} 1981), and a smaller one in the \( \pi^0 \pi^0 \) state (Edwards \textit{et al} 1982; Brandelik \textit{et al} 1981), and these features require a modest model dependence of the corresponding data analysis. Similarly the \( A_2 \rightarrow \gamma \gamma \) couplings have been measured through both the decay modes \( \eta \pi^0 \) (Edwards \textit{et al} 1982; Weinstein \textit{et al} 1983), and \( pp \) (Behrend \textit{et al} 1982, 1983), again with the helicity-two assumption, but with suitable parametrizations of the (modest) background effects. Similar remarks apply to \( f^0 \rightarrow \gamma \gamma \).

### \( P \rightarrow \gamma \gamma \) couplings

The traditional example of the \( P \rightarrow \gamma \gamma \) coupling is \( \gamma \gamma \) decay of a \( \pi^0 \) which has not only played a central role in establishing the \textit{colour} degree of freedom of quarks, but also acted as a laboratory for testing the twin concepts of \textit{pcac} and the axial vector anomaly associated with the so-called triangle diagram (Adler 1969; Bell 1969). In this respect we recall the original reasoning of Adler (1969) that the triangle (anomaly) diagram provides the “driving” matrix element for determining the \( \pi^0 \rightarrow \gamma \gamma \) amplitude by interpreting it as the vacuum \( \rightarrow \gamma \gamma \) transition of the \textit{pcac} term \( m_n^2 \phi_n \) in the anomaly-modified relation (Adler 1969; Bell 1969) for \( \partial_m j_n^\gamma \) and demanding that the latter should vanish in the limit \((k_1 + k_2)^2 = 0\). Thus the Alder reasoning provides a much-needed handle on the calculation of \( \pi^0 \gamma \gamma \) coupling at the physical \( \pi^0 \) mass \((k_1 + k_2)^2 = -m_n^2\), in terms of the Ward identity requirement that the \( \pi^0 \rightarrow \gamma \gamma \) amplitude should vanish at \((k_1 + k_2)^2 = 0\), but such a handle presupposes that \( \pi^0 \) is an \textit{elementary} field (Goldstone boson). On the other hand, in the standard quark-picture, the pion is just as much of a \( q \bar{q} \) \textit{composite} as any other meson, and figures 1 (b, c) tell precisely how the \( \pi^0 \) as a composite particle couples to \( \gamma \gamma \), without any further need to invoke the anomaly-modified \textit{pcac} relation to define its \( \gamma \gamma \) coupling at the physical pion mass (Adler 1969; Bell 1969). In other words figure 1 (b, c) not only play the roles of the anomaly diagrams themselves for the composite pion, but they also provide a complete description for the \( \pi^0 \rightarrow \gamma \gamma \) amplitude without any further assumptions (Adler 1969; Bell 1969). Exactly similar considerations hold for the \( \gamma \gamma \) couplings of the other pseudoscalar mesons.*

* This point of view is somewhat at variance with the position adopted in our earlier paper (Kulshreshtha and Mitra 1983) which we no longer consider tenable. The integration technique presented here for \( P \rightarrow \gamma \gamma \) is also numerically more accurate than the one employed earlier.
It is against this background that a comparison of the predictions with the data on $P \rightarrow \gamma\gamma$ couplings should be viewed. First of all we notice from (37) to (40) that the $M$-dependence of $\Gamma$ is $M^{13/4}$, roughly in accord with the phenomenological analysis (Kolankski 1984) based on $\Gamma \sim M^3$ (the balance of $M$-dependence being a dynamical effect). The predicted $\pi^0 \rightarrow \gamma\gamma$ and $\eta' \rightarrow \gamma\gamma$ rates, the latter on the basis of the mixing angle (22), are seen to be in good accord with the data, but the $\eta$ case seems, prima facie, to be a disaster. This is particularly so when it is remembered that the rate (0.56) estimated by the Crystal Ball group (Edwards et al 1982; Weinstein et al 1983) based on the analysis of the reaction

$$e^+ e^- \rightarrow e^+ e^- \eta \rightarrow e^+ e^- \gamma\gamma$$  (57)

is even higher than the more conventional value (0.324) quoted by PDG (1984) based on the Primakoff effect. This is a non-trivial discrepancy but a major part of it must be attributed to the $\eta$-$\eta'$ mixing effect. Indeed as seen from the mixing matrix, (21), the interference is constructive (and hence relatively insensitive) for $\eta' \rightarrow \gamma\gamma$. On the other hand, it is destructive for $\eta \rightarrow \gamma\gamma$, and hence highly sensitive to the $\theta$-value which is far from ‘ideal’ in this pseudoscalar case. (For further discussion, see Kolankski 1984)

As for the heavy quarkonium states $\eta_c, \eta_b$, unfortunately the results of direct measurements of the type (57) are not yet available. Rather, the data (Bloom et al 1983) listed in table 1 are highly indirect, being based on the observation (Kolankski 1984) that the wave function at the origin should be the same for both $1S_0(\eta_c$) and $3S_1(\psi/J)$ states, so that the accurately measured $e^+ e^- \gamma\gamma$ widths of the latter can be used to “determine” the corresponding $P \rightarrow \gamma\gamma$ widths through an essentially geometrical relationship. On the other hand it is possible to question the validity of this assumption, and indeed an independent estimate of $\eta_c \rightarrow \gamma\gamma$ based on the branching ratio for $\psi \rightarrow \gamma\eta_c$ yields a much smaller value, viz (Kolankski 1984)

$$\Gamma(\eta_c \rightarrow \gamma\gamma) = 1.6 \pm 0.8 \text{ keV},$$  (58)

which seems to be consistent with our theoretical prediction. (The rate for $\eta_b \rightarrow \gamma\gamma$ has been listed only for illustration, since in any case the $b\bar{b}$ model for harmonic confinement cannot be expected to work for $b\bar{b}$ spectroscopy).

de$\rightarrow gg$ width

We have attempted another independent check on the $\eta_c$ wave function through a prediction of the $\eta_c \rightarrow gg$ width as an $o\bar{z}$ forbidden process. The $\eta_c \rightarrow gg$ coupling can be trivially calculated in our model by making the following replacements:

\[ \text{Color: } \frac{1}{\sqrt{3}} \Rightarrow \frac{1}{\sqrt{3}} \delta_{ii}, \quad \frac{1}{\sqrt{3}} \text{Tr} \left( \frac{1}{2} \lambda^a \lambda^b \right) \Rightarrow \sqrt{8} \delta_{ab} = \sqrt{(2/3)}, \]  (59)

\[ \text{Charge: } \frac{e^2}{3\sqrt{2}} f_c \Rightarrow 4\pi a_s, \quad \text{where } \langle a_s \rangle \text{ is the "strong" coupling constant. Thus the overall replacement in (40) for estimating } \Gamma(P \rightarrow gg) \text{ becomes} \]  \[ \frac{\alpha}{\sqrt{6}} \Rightarrow \sqrt{(2/3)} \langle a_s \rangle. \]  (61)
bs dynamics for two-photon process

For a realistic estimate of $\langle \alpha_s \rangle$ appropriate to the confinement region, the following replacement of the perturbative $Q - \bar{Q}$ interaction ($= \alpha_s/r$) by the harmonic interaction

$$\frac{\alpha_s(r)}{r} = \frac{3}{4} \beta^2 \langle \bar{Q}Q \rangle r^2 = \frac{3}{2} m_c \omega^2 r^2$$  \hspace{1cm} (62)

should be in order, thus leading to the determination

$$\langle \alpha_s \rangle = \frac{3}{2} m_c \omega^2 \langle r^3 \rangle$$  \hspace{1cm} (63)

where $\langle r^3 \rangle$ must be calculated for the $Q\bar{Q}$ state under study. This gives

$$\langle \alpha_s \rangle = 3m_c \omega^2 / \beta^3 \sqrt{\pi}.$$  \hspace{1cm} (64)

Substitution of (61) and (64) in (40) yields the not unreasonable estimate

$$\Gamma(\eta_c \to gg) = 4.66 \text{ MeV},$$  \hspace{1cm} (65)

which compares favourably with the values $(11.5 \pm 5; 8.3 \pm 5)$ quoted by Bloom et al (1983). On the other hand, the same formula (62) yields for a typical ozi allowed process the value

$$\Gamma(\rho^0 \to gg) = 45.95 \text{ MeV}.$$  \hspace{1cm} (66)

which is only about 25\% of the total $\rho^0$ width (PDG 1984).

To summarise, the overall consistency of the several $T \to \gamma\gamma$ data with the predictions should warrant a cautious optimism about the dynamical status of the $q\bar{q}$ structure of these hadrons in the bs model for harmonic confinement. Notwithstanding this conclusion, it need hardly be emphasized that such a limited comparison of the data at $H \to \gamma\gamma$ coupling level, requiring as it does a model-dependent data analysis, cannot be a substitute for a more direct comparison with the "raw" data available in the form of various two-photon inelastic cross-sections themselves. Such calculations are envisaged in the near future.

Acknowledgement

We have enjoyed interesting discussions with Anuradha Mittal.

References

Adler S L 1969 Phys. Rev. 177 2426
Bell J S and Jackiw R 1969 Nuovo Cimento A60 47
Dainton J 1983 in Proc. of EPS Inti. Conf. on HEP, Brighton (eds) J Guy and C Costain
Fronsdal C Blankenbecler R and Taylor J C quoted in Mitra and Sood 1977
Mitra A N and Santhanam I 1981a Zeits. F. Phys. C8 25, 33
Mitra A N and Mittal A 1984 Pramana 22 221
Mitra A N and Mittal A 1985 (in preparation)
Mitra A N and Sood S 1977 Fortschritte der Physik 25 649
Particle Data Group, 1984 Rev. Mod. Phys. 56 No. 2
Adiabatic time-dependent Hartree-Fock theory with the consistency condition

A K MUKHERJEE and M K PAL
Saha Institute of Nuclear Physics, Calcutta 700 009, India

Abstract. Adiabatic time-dependent Hartree-Fock (atdhf) theory, including the authors' work in this field, has been summarised. In response to the criticism of Yamamura et al the role of curvature in preventing the choice of pure rpa mode as the solution near the static Hartree-Fock minimum has been discussed.

Keywords. atdhf; nuclear fission; Lipkin model; pure rpa mode.

PACS No. 21·60; 25·85

1. Introduction

In recent years there has been an increasing interest among nuclear theorists in developing a microscopic theory for the description of nuclear dynamics. The success of the mean-field approximation in describing the static properties of nuclei as well as the incredible ease of the Hartree-Fock (hf) calculation (Vautherin and Brink 1973) attained with the use of the density-dependent effective two-body interaction suggests that the mean-field approximation could also be used to describe nuclear dynamics. The old formulation on the time-dependent Hartree-Fock (tdhf) theory due to Dirac (1930) has been revived (Bonche et al 1976) designing it for the description of the time-dependent nuclear processes, especially for the large amplitude collective motions of nuclei. Sufficient progress has been made (Negele 1982) in the last decade in applying the tdhf theory for the description of heavy-ion collision and encouraging results for the deep-inelastic scattering, fusion events etc have been obtained. However, the limitations of tdhf in describing such many-body processes like nuclear fission, for which a description in terms of barrier penetration is traditionally employed, have become clear (Reinhardt 1981), the tdhf solution gets trapped inside the barrier and can never lead to fission into separated fragments.

An independent formalism, namely, the adiabatic time-dependent Hartree-Fock (atdhf) theory has been formulated (Baranger and Veneroni 1978; Brink et al 1976; Villars 1977; Moya De Guerra and Villars 1977; Goeke and Reinhard 1978) in the last few years especially designed for those processes whose dynamical evolution is governed by one or a few collective variables and in which the collective motion is slow compared to the single particle oscillations. Typical examples of such process are the nuclear fission and the inverse process, the sub-barrier fusion of two heavy ions.

Investigations have been carried out by several groups of workers (Reinhard and Goeke 1978; Goeke et al 1981; Rowe 1982; Rowe and Ryman 1982; Moya De Guerra and Villars 1978; Mukherjee and Pal 1981, 1982a, b) on whether the conditions derived from the mean-field theory of the adiabatic many-body process provide a unique
specification of the collective subspace. It has turned out in these investigations (Mukherjee and Pal 1982a, b; Pal 1984) that the zeroth- and first-order ATDHF equations, called the Villars equations, supplemented with a consistency condition derived from the second-order ATDHF equation specify the collective path uniquely. Further, it is a satisfying result, as has been proved analytically by Mukherjee and Pal (1982a) that the path followed by the many-body system during its adiabatic motion is indeed the 'fission path' which follows the bottom of the valley. In §2 a short review of the ATDHF theory has been included.

However there has been a certain criticism due to Yamamura et al (1984) on the results obtained by Mukherjee and Pal (1981) regarding the non-uniqueness of the solution of Villars equations. In the context of the three-level Lipkin model it has been argued by Yamamura et al (1984) that in the neighbourhood of a critical point of the collective potential the canonicity condition along with the RPA boundary condition restricts the solutions of the Villars equations to be unique and is given by the lowest-frequency principal line. In §3 emphasizing the role of curvature of the collective path, the ambiguity of the prescription used by Yamamura et al (1984) in order to choose the unique path has been clarified. Concluding remarks are presented in §4.

2. The ATDHF theory

The ATDHF theory has been developed primarily by Baranger and Veneroni (bv) (1978), Brink et al (1976), Villars (1977), Moya De Guerra and Villars (1977).

Baranger and Veneroni (1978) used the language of density matrix and the theory was cast in a form independent of any initial parametrisation. Using the TDHF equation the equations of motion as a function of time for the evolution of the time-even single-particle density matrix and the corresponding momentum matrix have been obtained; adiabaticity in this work implies that the momentum matrix is 'small' in some sense. Given the initial values of the time-even density and the momentum matrix the equations of motion can be integrated taking small time steps. Using the density and the momentum matrix the existence of a collective Hamiltonian has been established and the Schrödinger equation corresponding to the classical Hamiltonian can be solved to compute any observables pertaining to the collective motion.

In the formulation of Villars and coworkers (Villars 1977; Moya De Guerra and Villars 1977) the theory is cast in terms of the state vectors that are parametrised right from the beginning in terms of two parameters—one to play the role of collective coordinate $q$ and the other that of the collective momentum $p$. Assuming that the momentum $p$ is small which assures the adiabaticity the TDHF equation has been used (Villars 1977) perturbatively in powers of the adiabaticity parameter $p$ and the hierarchy of ATDHF equations has been obtained (Villars 1977). The zeroth- and first-order ATDHF equations, usually called the Villars equations, are given by

$$\langle \delta \Phi(q) \left| H - \frac{dV}{dq} Q \right| \Phi(q) \rangle = 0, \quad \text{(I)}$$

$$\langle \delta \Phi(q) \left[ [H, iQ] - P/m(q) \right] \Phi(q) \rangle = 0. \quad \text{(II)}$$

Here $P$ and $Q$ are one-body particle-hole (p-h) type operators and the former is the generator of displacement in the time-even Slater determinantal space $|\Phi(q)\rangle$. The
collective potential $V(q)$ and the collective mass $m(q)$ are given by,

$$V(q) = \langle \Phi(q) | H | \Phi(q) \rangle,$$

$$m^{-1}(q) = \langle \Phi(q) | [ [H, iQ], iQ ] | \Phi(q) \rangle.$$

At any point in the Slater determinantal space defined by $| \Phi(q) \rangle$ and its $1p-1h$ excited states $| \Phi_{ph}(q) \rangle$, (I) is an equation for the determination of the $ph$-components of $Q$ in terms of those of the given Hamiltonian, while (II) determines the $ph$-components of the operator $P$ which is the generator of displacement. In obtaining the time-even state $| \Phi(q + \delta q) \rangle$ at the neighbouring point with the help of the generator $P$ a normalisation factor $(\lambda/m)$ is required which is obtained by choosing the collective submanifold (in this case a single collective co-ordinate $q$) in terms of a measuring operator $D$,

$$q = \langle \Phi(q) | D | \Phi(q) \rangle,$$

or

$$\langle \Phi(q) | [ D, -iP ] | \Phi(q) \rangle = 1.$$

It is to be emphasized here that the direction in the $ph$-space defined by the operator $P$ is independent of the choice of this normalisation.

In this manner starting from any point (not a critical point) in the Slater determinantal space a unique path can always be obtained. By varying the starting point one can thus generate an infinite number of non-intersecting paths which are solutions of the Villars equations and converge at the nodal point of the potential function. It should be noted here that at the critical points where $dV/dq = 0$ (I) does not determine $Q$ and hence the procedure for generating a solution of the Villars equations (I) and (II), as described above, does not go through if a critical point is chosen in practice to start the solution.

In order to explore the feasibility of obtaining the ATDHF path, the theory has been applied (Moya De Guerra and Villars 1978; Mukherjee and Pal 1981) to describe a solvable model, namely, the three-level Lipkin model. In the context of the three-level model the Villars equations reduce (Moya De Guerra and Villars 1978; Mukherjee and Pal 1981) to a non-linear first-order ordinary differential equation of which the critical points of the collective potential i.e. the minimum, maximum, the saddle point etc are the singular points. All the solutions of the differential equation passes through the HF minimum which is then called the nodal point. With the use of a property regarding the behaviour of the solutions of the differential equation near the nodal point it has been shown (Mukherjee and Pal 1981) analytically that at the HF minimum the slopes of all the solutions including the valley path but excluding the higher-frequency principal line are the same as that of the lowest frequency $rpa$ mode and thus the non-uniqueness of the solution of the Villars equations has been established.

The solutions of Villars equations are shown (Mukherjee and Pal 1982a) to represent the lines of force orthogonal to the equipotential surface on the curved space of the time-even Slater determinants and thus the non-uniqueness of the solution has been established in general, irrespective of the model. Through any point in the space of time-even Slater determinants there exists a line of force representing the solution of the Villars equations at that point. All these infinite numbers of non-intersecting solutions converge at the HF minimum and therefore the Villars equations with the boundary condition that the solution should behave like the lowest-frequency $rpa$ mode near the HF minimum does not in general yield a unique ATDHF path. In particular in the case of a finite dimensional model like the Lipkin model where there are only two $rpa$ modes, it is
obvious that an infinite number of lines of force representing the solution of the Villars equations ought to behave at the nodal point as either of the RPA modes.

It is to be noted here that in the \textit{ATDHF} theory one is specifically interested in extracting a collective submanifold (in this case a one-dimensional path) on which the dynamics of the many-body system is governed by the Hamilton's equations of motion corresponding to a classical collective Hamiltonian defined on the collective submanifold. It is, however, intuitive on physical grounds as has been recognized frequently in the past in the macroscopic description of nuclear fission (Hill and Wheeler 1953) that physical systems describing adiabatic motion should follow the path along which the energy is minimised \textit{i.e.} the path following the bottom of the valley. The valley of a potential function $V$ is defined variationally as,

\begin{equation}
\delta \left\{ |\text{grad } V|^2 \right\} - \Lambda \delta V = 0,
\end{equation}

\textit{i.e.} the magnitude of the potential gradient is stationary for excursions along the equipotential surfaces. Here $\Lambda$ plays the role of a Lagrange multiplier taking care of the constraint.

In view of the above non-uniqueness of the solution of Villars equations and the desire to explore the valley path as the optimal \textit{ATDHF} solution Goeke and coworkers (Goeke and Reinhard 1978; Reinhard and Goeke 1978; Goeke \textit{et al} 1980, 1981, 1983a, b) proposed a validity criteria for exploring the \textit{ATDHF} path. It has been argued that to make the adiabatic assumption valid, a second-order quantity obtained from the second-order \textit{ATDHF} equation should be \textit{small} compared to the first-order contribution. In practice, however, they find (Goeke \textit{et al} 1980, 1983a, b) that the ratio of the two contribution be \textit{minimum} along the optimal path. Indeed the validity of the adiabatic assumption requires that the second-order quantity should be \textit{small} compared to the first-order contribution; it is however not clear from theoretical grounds why their ratio should be \textit{minimum} for the optimal solution of the \textit{ATDHF} theory. Nevertheless in numerical calculations (Goeke \textit{et al} 1980, 1983a, b) they have used ‘trial and error’ method for exploring the optimal \textit{ATDHF} path.

However, to describe nuclear fission, the theory should provide in a consistent manner the optimal path as the lowest valley of the many-body potential energy surface. It has been established by Mukherjee and Pal (1982a) that simultaneous solution of the zeroth-, first- and second-order \textit{ATDHF} equations indeed delineates the lowest valley as the optimal solution of the \textit{ATDHF} theory.

The formal identity of the \textit{ATDHF} approaches of Villars and of Baranger and Veneroni (\textit{BV}) has been established (Mukherjee and Pal 1982a; Pal 1984). It has been demonstrated that while the \textit{BV} approach mixes the zeroth- and second-order quantities, Villars neglects the second-order quantity altogether. In the spirit of the perturbation approach the second-order equation,

\begin{equation}
\langle \delta \Phi(q) | \frac{1}{2} [ [H, iQ] , iQ ] - \frac{1}{2} \frac{d m^{-1}}{dq} Q + m^{-1} \frac{d Q}{dq} | \Phi(q) \rangle = 0,
\end{equation}

should be considered as a separate equation. It has been shown by Mukherjee and Pal (1982a) that for consistent exploitation of the time-dependent equation in the adiabatic limit the second-order equation (III) should be considered as a separate equation along with the zeroth- and first-order Villars equations (I) and (II).

Since, as has been noted earlier, the zeroth- and first-order equations can be solved,
given a time-even determinantal state as the initial condition, the simultaneous fulfilment of the second-order equation calls for a consistency condition,

\[ \langle \delta \Phi(q) | \frac{1}{2} \lambda^2 \left[[H, iQ], iQ\right] + \frac{\lambda}{m} [H, -iP] - \frac{1}{2} \lambda \omega_0(q) Q | \Phi(q) \rangle = 0 \quad (\text{III'}) \]

derived from the zeroth- and second-order equations to be satisfied. Here,

\[ \omega_0(q) = \frac{1}{\lambda} \frac{d}{dq} \left( \frac{\lambda^2}{m} \right). \quad (4) \]

It is a satisfying result as has been proved by Mukherjee and Pal (1982a) by an analytic method (without invoking any model) that the consistent solution of Villars equations (I) and (II) satisfying the consistency condition (III') yields a unique path. In the space of the time-even Slater determinant for which the mass matrix plays the role of the metric tensor characterising the space, the solution represents the valley of the collective potential satisfying the variational equation (3) with \( \omega_0(q) \) playing the role of the Lagrange multiplier \( \Lambda \).

It should be noted here that having obtained the optimal collective path, the classical Hamiltonian defined on the optimal path can be established and with the use of the Schrödinger equation corresponding to the classical Hamiltonian the observables pertaining to fission may be computed.

However it is to be noted that the \textit{ATDHF} theory indeed yields the Hamilton's equation of motion corresponding to the classical Hamiltonian,

\[ \mathcal{H}(p, q) = \frac{p^2}{2m(q)} + V(q) \quad (5) \]

provided the canonicity condition,

\[ \langle \Phi(q) | [Q, -iP] | \Phi(q) \rangle = 1 \quad (6) \]

is satisfied.

The canonicity condition (6) is interpreted (Villars 1977; Goeke and Reinhard 1978; Reinhard and Goeke 1978; Mukherjee and Pal 1982a, b; Goeke \textit{et al} 1980, 1983a, b) as a normalisation condition of the operators \( Q \) and \( P \). Along with the equation (2b) that specifies the collective co-ordinates in terms of the measuring operator \( D \), the canonicity condition (6) has been shown (Mukherjee and Pal 1982b; Goeke \textit{et al} 1980, 1983a, b) to yield an equation for determining the collective mass \( m(q) \) consistent with the definition (1b).

### 3. Non-uniqueness in the Lipkin model

In the context of the three-level Lipkin model Yamamura \textit{et al} (1984) obtained the solution of the linearised Villars equations (I) and (II) in the neighbourhood of a critical point which are in essence the same as those obtained by Mukherjee and Pal (1981). Yamamura \textit{et al} (1984) concluded that the slopes of all these solutions at the critical point are the same as that of the lowest frequency \textit{RPA} mode. The above conclusion is true only when the critical point is a nodal point characterised by the fact that the square of the \textit{RPA} eigenvalues are of the same sign. If the \textit{RPA} eigenvalues (squared) are of
opposite sign characterising the saddle point the principal lines which designate the RPA
dmode is the only solution (Mukherjee and Pal 1981).
The requirement of the RPA boundary condition demands that,
\( (\partial^2 V/\partial q^2)_{q=\text{HF}} = \lambda_1, \)
where \( \lambda_1 \) is the lowest eigenvalue of the RPA matrix.
The canonicity condition of (6) linearised about the critical point along with the
substitution of (7) and the afore-mentioned solution of the linearised Villars equations,
can be written, following Yamamura et al (1984) as
\( (A^2 - 1)\lambda_1^2 \xi^2 + B^2 \lambda_2^2 \xi^2 \mu = 0. \)
Here \( A \) and \( B \) are constants of integration, \( \lambda_1, \lambda_2 \) are the RPA eigenvalues (squared)
with \( \mu = \lambda_2/\lambda_1 \). In order to avoid confusion with the operator \( Q \) a new notation \( \xi = q - q_{\text{HF}} \) has been used in place of the notation \( Q \) of Yamamura et al (1984).
It is to be noted here that in deriving (8) the orthonormality of the RPA eigenvectors
with respect to the metric \( \mathbf{M}^{-1} \) (Mukherjee and Pal 1981) has been used.
It has been argued by Yamamura et al (1984) that (8) is satisfied if and only if,
\[ A^2 \approx 1, \]
\[ B = 0. \] (9b)
Condition (9) is a particular solution of the linearised Villars equations namely the
principal line designating the lowest-frequency RPA mode. Yamamura et al (1984)
concluded that the solution of the Villars equations (I) and (II) satisfying the canonicity
condition (Baranger and Veneroni 1978) yields a unique path which is given by the
solution of the ATDHF equations (I) and (II) satisfying the boundary condition that the
solution in the neighbourhood of the critical point is given by the lowest-frequency
principal line characterised by (9). The lacuna in this argument of Yamamura et al
(1984) will now be clarified.
Let us discuss the linearisation of the Villars equations (I) and (II) about the static
Hartree-Fock (HF) state. Since (I) is satisfied for all values of \( q \) its derivative with respect
to \( q \) should also be satisfied,
\[ \langle \delta \Phi(q) | [\mathbf{H}, - iP] - C(q)Q - \lambda(q) \frac{dQ}{dq} | \Phi(q) \rangle = 0. \]
Here \( \lambda(q) = dV/dq \) and \( C(q) = d\lambda/dq \). At the critical point of the potential where \( \lambda(q) = 0 \) (I) does not determine, as noted earlier after (2b), the p-h components of \( Q \); instead at that point it represents the static HF condition. In the neighbourhood of the HF point if the curvature of the path is neglected then \( dQ/dq = 0 \) and (10) can be written as,
\[ \langle \delta \Phi(q) | [\mathbf{H}, - iP] - C(q)Q | \Phi(q) \rangle = 0. \]
Note that exactly at the HF point, \( \lambda(q) = 0 \), the third term in (10) drops out and (11) is
exact.
Equation (11) along with (II) can easily be reduced to the RPA eigenvalue equation,
\[ \mathbf{R} | X_{\text{hp}} \rangle = \omega(q) | X_{\text{hp}} \rangle, \]
where \( \omega(q) = C(q)/m(q) \) and \( \mathbf{R} \) is the RPA matrix. We thus conclude that at the HF point
the solutions are the RPA modes and the lowest energy solution is the lowest RPA mode.
However, due to the curvature effect, this does not preclude the possibility of many solutions in the neighbourhood of the HF point terminating at that point with the slope of the pure lowest RPA mode.

The curvature term \( (dQ/dq) \) in (10) cannot be neglected in the neighbourhood of the HF point as can be seen from the second-order equation (III), which gives (Mukherjee and Pal 1982a)

\[
\langle \delta \Phi(q) \frac{dQ}{dq} | \Phi(q) \rangle = \frac{m}{2} \langle \delta \Phi(q) | [H, iQ], iQ] - \frac{dm^{-1}}{dq} Q | \Phi(q) \rangle. \tag{13}
\]

It therefore follows that although the solution of the ATDHF equations in the limit \( (q \to q_{HF}) \) behaves like the decoupled RPA solution, in the neighbourhood of the HF minimum, the solution cannot be given by a pure RPA mode but is a superposition of the RPA modes which is dictated by the non-vanishing curvature.

The above assertion has been demonstrated in the context of the three-level Lipkin model by Mukherjee and Pal (1981) and in order to clarify further we reproduce here (figure 1) one of the figures of Mukherjee and Pal (1981) where the lowest-frequency RPA mode is shown as the line HR which excepting at the HF point is not an ATDHF solution. In the ATDHF numerical calculations (Goeke et al 1980, 1983a, b) of the fusion of \( ^{16}O + ^{16}O \) this ambiguity of starting the solution at the HF state of the compound system has been clearly felt and the optimal path is obtained by trial and error method starting from a point far off from the HF state.

![Figure 1](image-url)

*Figure 1.* The entire topography of ATDHF paths (continuous lines) along with the contour plot (dashed lines) of \( V(\theta_1, \theta_2) \) for the case \( x > (n+1) \) \( (x = 3, n = 1) \) with the maximum \( M \), minimum \( H \) and the two saddle points \( S_1 \) and \( S_2 \) of \( V(\theta_1, \theta_2) \) are shown. The optimal path following the bottom of the valley is shown by heavy dots. The line HR (parallel to the \( \theta_2 \)-axis) represents the lowest-frequency RPA mode.
In the adiabatic motion the nuclear system will in general start oscillations in a superposed mode and is expected to choose an optimal path along which the energy is minimised. This is the path which satisfies the consistency condition (III'). We therefore conclude that instead of constraining the system to choose a pure \textit{RPA} mode, the system should be given the freedom to choose its lowest energy path which is obtained numerically by satisfying the canonicity condition. Constraining the solution deliberately as is done in Yamamura \textit{et al.} (1984) may be at best a good mathematical exercise with no utility for the physics of the large-amplitude nuclear collective motion for which delineating the valley path is to be set as a physically more desirable goal. Even within the three-level model the solution generated with the mathematical constraint of Yamamura \textit{et al.} (1984) departs considerably from the valley path as seen from the figure.

In the context of the three-level Lipkin model it thus follows from the above discussions that even in the neighbourhood of the HF point, the principal lines are not a solution to the \textit{ATDFH} equations (I) and (II) which in the Lipkin model reduce to a non-linear first-order ordinary differential equation and thus the prescription of solving (I) and (II) or equivalently the non-linear differential equation with a boundary condition of particular solution (in the neighbourhood of the HF point) which itself is not a solution to the \textit{ATDFH} equations is not clear to the present authors.

In our view (8) does not always imply that the condition (9) should be satisfied because of the following reason. The solutions of Villars equations (I) and (II) are obtained by linearising (I) about the critical point and therefore in essence is equivalent to (11) in which the curvature has been neglected. Further the canonicity condition is also linearised about the critical point. In such a linearised \textit{ATDFH} theory we can only demand that (8) should be satisfied only in the limit \((\xi \to 0)\). If the limit \((\xi \to 0)\) of (8) is utilised two situations arise:

(i) If \(\mu > 0\) characterising the nodal point then in the limit \((\xi \to 0)\) (8) is satisfied for \textit{all values} of \(A\) and \(B\) including those of (9) and thus the principal line is not the only solution that satisfies the canonicity condition (8) but a whole host of solutions of Villars equations along with the principal line satisfies the canonicity condition and behave like the lowest-frequency \textit{RPA} mode near the HF minimum.

(ii) If \(\mu < 0\) characterising the saddle point then in the limit \((\xi \to 0)\) (8) implies that (9) must be satisfied and thus the principal line is the only solution.

Thus the assertion of Yamamura \textit{et al.} (1984) that in the neighbourhood of a \textit{critical point} the only solution of Villars equations that satisfies the canonicity condition (6) and \textit{behave} like the lowest frequency \textit{RPA} mode is the lowest-frequency principal line is true only when the \textit{critical point} is a \textit{saddle point} but \textit{not a nodal point} and therefore the results established in our work (Mukherjee and Pal 1981, 1982a, b) contrary to the criticism of Yamamura \textit{et al.} (1984) are correct and valid.

4. Conclusions

In this paper the non-uniqueness of the solution of Villars equations satisfying the canonicity equation and the \textit{RPA} boundary condition, in contrast to the criticism due to Yamamura \textit{et al.} (1984) has been established. The adiabatic time-dependent Hartree-Fock theory with the consistency condition which delineates in a consistent manner the 'fission path' desired by workers in this field has been summarised.
Until now the application of the ATDHF theory to realistic calculations has been limited to the fusion of two heavy-ions; the work has been carried out by Goeke et al (1980, 1983a, b). However the ATDHF theory is most suitable for describing nuclear fission as is clear from the text. So far it has not been applied to nuclear fission. Though theoretical investigations are required to incorporate some of the relevant features of fission, the numerical application of the theory to nuclear fission is the challenging problem to nuclear theorists in recent times.

Acknowledgement

One of the authors (AKM) wishes to thank D Shyam for useful discussions.

References

Dirac P A M 1930 Proc. Cambridge Philos. Soc. 26 376
Hill D L and Wheeler J A 1953 Phys. Rev. 89 1102
Marumori T 1973 Prog. Theor. Phys. 57 112
Negele J W 1982 Rev. Mod. Phys. 54 913
Pal M K 1984 in Proc. of the Winter College of Fundamental Nuclear Physics, Trieste, Italy, 7th Feb, 30th March 1984
Reinhard P G and Goeke K 1978 Nucl. Phys. A312 121
Rowe D J and Basserman R 1976 Can. J. Phys. 54 1941
Vautherin D and Brink D M 1973 Phys. Rev. C5 626
Nucleon transport processes in fission and heavy ion reactions

S S KAPOOR and V S RAMAMURTHY
Bhabha Atomic Research Centre, Bombay 400085, India

Abstract. The nucleon exchange process between two nuclei in close proximity and its application to an explanation of fragment mass and charge distributions in fission and in heavy ion deep inelastic collisions are reviewed. An analysis of the measured correlations between the energy loss from relative motion and the fragment mass and charge variances in the heavy ion deep inelastic collisions is presented. The recent data on fragment mass and charge variances as a function of the fragment kinetic energy in thermal neutron induced fission of $^{235}$U, lends added support to the hypothesis that the nucleon transport process plays a similar role both in fission and in heavy ion deep inelastic collisions.

Keywords. Nucleon transport process; fission; heavy ion reaction; stochastic theory; nucleus-nucleus collision.

PACS No. 25:70; 25:85

1. Introduction

Nucleon exchange between the two nascent fragments formed during the fission process near the scission point as the mechanism leading to the observed mass distributions in fission was first proposed by Ramanna (1964) nearly two decades ago. More detailed models (Ramanna et al 1965; Ramanna and Ramamurthy 1969; Ramamurthy 1971; Prakash et al 1980) based on the above nucleon exchange mechanism were subsequently developed to satisfactorily explain the general features of the fragment mass and charge distributions in fission. In recent years, with the availability of a variety of accelerated heavy ion beams to study heavy ion-induced nuclear reactions, the nucleon exchange process taking place between two nuclei in close proximity has found additional experimental support from the study of deep inelastic heavy ion collisions. In fact, studies of nucleus-nucleus collisions at medium energies have now made possible investigations of the nucleon transport processes in a systematic and almost controlled manner. During the last decade, the large body of experimental data on the conversion of the relative kinetic energy into fragment excitation energies (commonly referred to as energy dissipation or energy loss), transfer of orbital angular momentum to the fragment spins and nucleon transfers between the target and the projectile (see reviews by Schroder and Huizenga 1977; Lefort and Ngo 1978; Weiden Muller 1980; Gobbi and Norenberg 1980; Kapoor 1982 and references cited therein). These studies have also established that a significant number of nucleon exchanges take place in the very short interaction
times of few times $10^{-21}$ sec involved in the deep inelastic collisions. In this work, we present a brief review of the nucleon exchange process operating between two nuclei in proximity, and its application to the fission process and heavy ion reactions.

2. Stochastic theory of fragment mass and charge distributions in fission

The application of the nucleon exchange mechanism for an explanation of the fragment mass and charge distributions in fission is based on the fact that during the last stages of the fission process, the fissioning nucleus can be well approximated by two weakly interacting nascent fragments. If one assumes that the relative motion of the nascent fragments is slow compared to the characteristic time for a nucleon transfer, the process can be treated as a stochastic process where not only the nascent fragments are in thermodynamic equilibrium within themselves but they also attain mutual equilibrium among themselves with respect to nucleon and energy transfers. The observed distributions of the fragment mass and charge then correspond to the equilibrium distribution near the scission point. Ramanna and coworkers (Ramanna et al. 1965; Ramanna and Ramamurthy 1969) have given the theoretical formulation of the nucleon exchange process between the two nascent fragments for the calculation of the equilibrium mass and charge distributions. Considering for the sake of illustration a one-component system, the configuration of the fissioning nucleus at any instant near the scission point can be defined by specifying the number of nucleons on either side. Let $\omega_M$ be the probability that the fissioning nucleus has a configuration with $M$ nucleons in the heavy side, the number of nucleons in the light side being determined from the total nucleon number conservation. If $P_{M',M}$ denotes the probability that a configuration with $M'$ nucleons in the heavy side goes over to the configuration with $M$ nucleons in the same side in a small interval of time $\Delta t$, one has

$$w_M(t + \Delta t) = \sum_{M'} w_{M'}(t) P_{M',M}$$

By definition

$$\sum_M P_{M',M} = 1 \quad \text{and} \quad \sum_M w_M = 1$$

Under the condition of complete equilibrium in the mass asymmetry degree of freedom, one has $w_M(t + \Delta t) = w_M(t)$. If the unit of time $\Delta t$ is sufficiently small and one can neglect cluster transfers

$$P_{M',M} = 0, \quad M \neq M' \text{ or } M' \pm 1$$

It has been shown that under these conditions the probabilities $w_M$ follow the simple relation

$$\frac{w_{M+1}}{w_M} = \frac{P_{M,M+1}}{P_{M+1,M}}$$

(1)

Thus the ratio of the probabilities for adjacent configurations $M$ and $M + 1$ are simply equal to the ratio of the nucleon transfer probabilities in the directions $M \rightarrow M + 1$ and $M + 1 \rightarrow M$. A generalization of (1) in two dimensions to consider both proton and neutron transfers, gives

$$w_{N,Z}(t + \Delta t) = \sum_{N'} \sum_{Z'} w_{N',Z'}(t) P_{N'Z',NZ}$$

(2)
The steady state solution of $w$ is given by

$$w_{N,Z} = \sum_{N'} \sum_{Z'} w_{N',Z'} P_{N',Z'}$$

Equation (3) can be solved numerically to obtain the probability distribution $w$, provided the transition probabilities are known. It is seen from the above equations that the central quantities which decide the mass and the charge distributions are the transition probabilities $P_{N,Z}$ which are in turn given by the single-nucleon transfer probabilities from one of the fragments to the other.

Figure 1 illustrates the mechanism of nucleon transfers between the two nascent fragments. If the fragments are cold, the direction of spontaneous transfer of nucleons is from the fragment having the higher Fermi energy to the one having lower Fermi energy. Even if the nascent fragments have some excitation energy, a tendency for a preferential transfer of nucleons in the direction of decreasing chemical potential.

Figure 1. Schematic diagram illustrating the nucleon exchange mechanism between the fragments $L$ and $H$. $\rho_{L/H}(r)$ represents the diffuse densities, $g_{L/H}$ are the single particle level densities, $f_{L/H}$ are the Fermi-Dirac occupation probabilities and $\mu_{L/H}$ are the chemical potentials of the fragments $L$ and $H$ respectively.
persists. A quantitative estimate of the relative probabilities of nucleon transfers in both directions can be made as (Prakash et al 1980; Ramanna and Ramamurthy 1969)

\[
P_{M,M+1} = \frac{2\pi}{\hbar} \int g_L(E) f_L(E) g_H(E) [1 - f_H(E)] \left| M_{LH} \right|^2 dE
\]

(7)

\[
P_{M,M-1} = \frac{2\pi}{\hbar} \int g_H(E) f_H(E) g_L(E) [1 - f_L(E)] \left| M_{HL} \right|^2 dE
\]

(8)

Here \(g_L(E)\) and \(g_H(E)\) are the single particle energy level densities and \(f_L(E)\) and \(f_H(E)\) are the Fermi-Dirac occupation probabilities in the light and the heavy nascent fragment respectively. The matrix elements of transfer \(M_{LH}\) and \(M_{HL}\) are equal because of microscopic reversibility. Since the main contribution to the integral comes from a small energy band around the chemical potentials of the two nascent fragments, the quantities \(g_L\), \(g_H\) and the matrix element of transfer can be calculated at the mean chemical potential and taken out of the integral. One then has

\[
P_{M,M+1} = \frac{2\pi}{\hbar} \left| \frac{\overline{M}}{2} \right|^2 g_L(\overline{\mu}) g_H(\overline{\mu}) I_{LH}
\]

(9)

where

\[
I_{LH} = \int f_L(E) [1 - f_H(E)] dE
\]

(10)

and \(\overline{\mu}\) is the mean chemical potential. A similar expression for \(P_{M,M-1}\) can also be written. For a nearly degenerate system like a nucleus having a temperature much less than the mean chemical potential, one gets

\[
I_{LH} = \frac{(\mu_H - \mu_L)}{e^{(\mu_H - \mu_L)/T} - 1}
\]

(11)

where \(\mu_H\) and \(\mu_L\) are the chemical potentials of the two nascent fragments. Thus the main driving force for the net transfer of nucleons from one of the fragments to the other is the difference in their chemical potentials though a finite temperature induces nucleon transfers in both directions.

In the limit of zero temperature, the difference in the chemical potentials is equal to the negative of the difference in the nucleon separation energies of the two nascent fragments, neglecting the small rearrangement energies. For finite temperatures, one can include a temperature dependence of the chemical potential in such a way that the shell effects on the chemical potential vanish at high temperatures. Figure 2 shows a typical plot of the difference in the chemical potentials versus the mass ratio for the case of \(^{236}\)U fission (Prakash et al 1980). The fact that the difference \((\mu_H - \mu_L)\) is negative for all configurations up to neutron number \(N = 86\) and is positive for all configurations above \(N = 86\) is responsible for the increasing yields of the heavy fragments up to \(N = 86\) and a decrease in the yield thereafter. Figure 3 shows typical mass distributions for the fission of \(^{226}\)Ra, \(^{252}\)Cf and \(^{256}\)Fm as calculated from this model. It can be seen that the calculations reproduce the known qualitative features for all cases and even such details as the triple hump for \(^{226}\)Ra fission are reproduced.

One of the early criticisms of the nucleon exchange model for explaining the fragment mass and charge distributions in fission was whether sufficient time is
available for an appreciable number of nucleon transfers to take place during the
descent from the saddle to the scission point. Direct experimental evidence for
substantial nucleon exchange between two nuclei in close proximity in interaction times
of the order of $10^{-21}$ sec has now come from the studies of heavy ion deep inelastic
collisions.

3. Nucleon exchange processes in nucleus-nucleus collisions

It is now well known that in reactions induced by heavy ions, a sizable cross-section
appears in a new process called deep inelastic collisions. These reactions are
dcharacterised by a rather short interaction time of few times $10^{-21}$ sec during which a
large fraction of the kinetic energy of the colliding nuclei gets converted into intrinsic
fragment excitation energies. In addition, a large number of nucleons are also
exchanged between the target and the projectile as inferred from the observed width of
the mass and charge distributions of the target-like and the projectile-like binary
fragments emitted as the reaction products. An important observation is the presence of
strong correlation between various quantities such as the kinetic energy loss from
relative motion, fragment mass and charge widths and the angular momentum transfer
(see for example, Wollersheim et al. 1982; Dyer et al. 1980; Dakowski et al. 1982). A
nucleon exchange mechanism similar to the one proposed earlier for an explanation of
fragment mass and charge distributions in fission has been found to be useful in
bringing out the main features of the deep inelastic collision process. One difference
between the earlier model for fission and that for deep inelastic collisions comes from
the effect of the relative motion of the two ions exchanging the nucleons. In addition,
unlike the case of fission, in deep inelastic collisions the number of nucleon transfers are
not sufficient to attain complete equilibration in the nucleon exchange degree of
freedom. Under these conditions, it is advantageous to define the nucleon drift and the
S S Kapoor and V S Ramamurthy

Figure 3. Calculated fragment mass yields for the fission of $^{226}$Ra, $^{252}$Cf and $^{256}$Fm with a mean fragment temperature of $T = 0.5$ MeV.

diffusion coefficients within the Fokker-Planck approximation (Prakash et al 1981). Theoretical formulations of the nucleon exchange process between two ions in relative motion and the resulting transport coefficients can be found in the recent literature pertaining to the heavy-ion deep inelastic collisions.

Considering the relative motion between the colliding nuclei, the fermion nature of the exchanged particles and the associated Pauli blocking effect, it has been shown (Randrup 1977) that the diffusion of nucleons between the two nuclei results in a time variation of the width of the fragment mass distribution as

$$\frac{d\sigma^2}{dt} = N_A' \langle w \coth \frac{w}{2T} \rangle_F$$

where $N_A'$ is the differential total particle current, $w$ is the change in excitation energy associated with the transfer of a nucleon, $T$ is the temperature of the system. The suffix $F$ denotes that the average is taken around the mean value of the Fermi energies of the two nuclei. It is easy to see that the quantity $w$, denoting the change in excitation energy on account of transfer of a nucleon from one nucleus to the other having a relative
velocity $v$ is given by

$$w = F_A - p \cdot v$$

where $F_A$ is the difference in the Fermi energies and $p$ is the momentum of the transferred nucleon.

The rate of the resulting energy dissipation is given by

$$- \frac{dE}{dt} = N_A' <w^2>_F$$

(13)

In most cases of the deep inelastic heavy ion collisions, it is a good approximation to take

$$\coth \left( \frac{<w^2>_F}{T} \right) \approx 1$$

and (12) is then often approximated as

$$d\sigma_A^2/dt \approx N_A' (<w^2>_F)^\frac{1}{2}$$

(14)

From (13) and (14) it then follows that

$$- dE/d\sigma_A^2 = (<w^2>_F)^\frac{1}{2}$$

(15)

For peripheral collisions, where most of the nucleon transfers take place, (15) reduces to

$$- dE/d\sigma_A^2 \approx \left[ \frac{m}{\mu} EE_F \right]^{\frac{1}{2}} \approx \alpha \frac{m}{\mu} E$$

(16)

where $m$ is the nucleon mass, $\mu$ is the reduced mass of the di-nuclear complex and $E_F$ is the average of the Fermi kinetic energies of the two nuclei. $E$ is the kinetic energy in relative motion above the Coulomb barrier $V_c$ given by

$$E = E_{cm} - V_c (R_{int}) - E_{loss}$$

$R_{int}$ refers to the interaction distance between the colliding ions, and $E_{loss}$ is the energy loss from the relative motion. In a slightly different approximation, one obtains

$$d\sigma_A^2/dt \approx N_A' <|w|>_F$$

which when flux-averaged gives an additional factor of $3\pi/8$ over that of (16). Although we base the following discussion on the use of (16), the main conclusions would remain unaltered if the factor $3\pi/8$ is included. For the classical case of exchanged particles being motionless in the individual nuclear containers, $\alpha$ will be unity. Thus, the value of

$$\alpha = \left( \frac{\mu E_F}{m E} \right)^\frac{1}{2}$$

in (16) results from the consideration of the Fermi nature of the exchanged particles and the Pauli blocking effect.

On integrating (16), one gets

$$E^{1/2} \approx E_0^{1/2} - \frac{1}{2} \left( \frac{m}{\mu} E_F \right)^{1/2} \sigma_A^2$$

(17)

The Fermi energy $E_F$ is a reasonably well-known quantity ($\sim 37$ MeV), and thus the above predictions of the nucleon exchange model do not involve any free parameter.
Before we discuss the comparison of the experimental results with the predictions of (17) some additional facts must be pointed out. In most experiments the quantity which is measured is $\sigma^2_z$, the variance of the fragment charge distribution and not $\sigma^2_x$. Conversion of $\sigma^2_x$ into $\sigma^2_z$ requires a knowledge of the neutron-proton correlation in the exchange process. For uncorrelated transfers $\sigma^2_x = (A/Z)\sigma^2_z$ while for fully correlated transfers $\sigma^2_x = (A/Z)^2 \sigma^2_z$. Thus, in some earlier studies the comparison of the correlation between the measured $\sigma^2_z$ and the energy loss with the predictions of the above transport model involved some ambiguity as a result of uncertainty in the conversion of $\sigma^2_z$ into $\sigma^2_x$ representing the total number of nucleon exchanges. Further studies of this problem carried out at Trombay (Kapoor and De 1982; De and Kapoor 1983) have clarified that it is possible to make comparisons of experimental results on $\sigma^2_z$ with the theoretical predictions of the transport model without much ambiguity arising from the degree of neutron-proton correlation. It was pointed out in this work that if there is a neutron-proton correlation in the exchange process, not only should the experimental $\sigma^2_z$ be suitably converted into $\sigma^2_x$ but one must also consider that the theoretical predictions of the transport model also get modified. It was shown (Kapoor and De 1982; De and Kapoor 1983) that while in the case of uncorrelated neutron-proton transfers, the variance of the fragment mass distribution $\sigma^2_x$ is equal to the total number of nucleon transfers, for fully correlated motion, the value of $\sigma^2_x$ is larger than the number of nucleon transfers by a factor $A^2/2ZN$. If one considers the correlated and uncorrelated motion for a given number of particle exchanges, the rate of energy dissipation remains unaltered. It then follows that the theoretical value of $dE/d\sigma^2_x$ for correlated transfer is related to the expression for uncorrelated transfers as

$$
\left[\frac{dE}{d\sigma^2_x}\right]_{\text{corr}} = \frac{2ZN}{A^2} \left[\frac{dE}{d\sigma^2_x}\right]_{\text{uncorr}}
$$

The experimental results on the variance of the fragment charge distributions versus energy loss for several heavy ion reactions were analysed earlier (Kapoor and De 1982; De and Kapoor 1983; Kapoor 1982a) by including the effect of correlations in the transport model in a manner which is consistent with the assumption made in the transformation of $\sigma^2_z$ into experimental $\sigma^2_x$. Figures 4 and 5 show the comparisons with the experimental results for cases of heavy ion reactions of Xe + Bi, Pb + U, and Pb + Pb. The dashed curve in figure 4 represents the theoretical curve if one is not consistent in applying the effect of correlation in neutron-proton exchanges in both the transport model, and in the relationship between $\sigma^2_z$ and experimental $\sigma^2_x$. The good fits of the experimental points to the solid lines in the figures show satisfactory agreement with model predictions, nearly independent of the assumptions made about neutron-proton correlations in the exchange process, bringing out unambiguously the dominant role of nucleon-exchange process in the energy loss.

The influence of the neutron-proton correlation in the evolution of the fragment mass and charge distributions in fission has been studied recently at Trombay (Rekha Govil et al 1983). Making use of a back-to-back $\Delta E - E$ detector arrangement, the correlation between the variances in the fragment mass and charge distributions versus the total fragment kinetic energy was studied in the case of thermal neutron induced fission of $^{235}$U. Figure 6 shows the measured correlations and shows a rather close resemblance to the known systematics (Breuer et al 1979) in the case of heavy ion deep inelastic collisions, namely, the ratio $\sigma^2_x/\sigma^2_z$ is close to $(A/Z)^2$ for low kinetic energies and approaches $(A/Z)$ for higher kinetic energies. This similarity lends support to the
hypothesis that the final mass and charge distributions in fission as well as in DIC are governed by similar processes, namely, a nuclear exchange process.

There has also been some discussion in the earlier work on the question of the nuclear shell effects on the nucleon exchange process. In the analysis of the heavy ion deep inelastic collision data with (17), the single particle states of the nucleons are assumed to be given by a smooth distribution, and therefore nuclear shell effects are neglected. Good fits obtained above for the cases of Bi + Xe and Pb + U do indicate that the shell effects have little influence on the above description which may partly be due to the washing out of the shell effects at finite temperatures reached during the nucleon exchange process. However, for the case of Pb + Pb, where both the colliding nuclei have doubly closed shells and where we are dealing with nuclei which have highest shell correction energy, noticeable deviations from the predictions of the Fermi gas model are apparent in figure 5. Although bulk of the energy loss in heavy ion collisions originates from the nucleon exchange process, it is also to be expected that some energy loss arises from other competing mechanisms. One effect of the energy gap at the Fermi-surface, that is, of shell effects, would be to slow down the nucleon exchange process, on account of a smaller density of states at the Fermi surface. This could then
relatively enhance the importance of other mechanisms competing with the nucleon exchange process, resulting in deviations of the theoretical predictions from the experimental results. However, this question of nuclear shell effects in the nucleon exchange process needs to be further investigated.

Thus, the main conclusions resulting from the above discussions are as follows:

When the effect of isospin correlations in the exchange process is considered both in the conversion of measured $\sigma_2^2$ into experimental $\sigma_2^1$, and also in the transport model, it turns out that comparison of $\sigma_2^1$ versus energy loss with the model is almost independent of the degree of the isospin correlation in the exchange process. Thus such a comparison focuses on the magnitude of the energy loss arising from the exchange process without much ambiguity due to the presence of isospin correlation. The analysis of a number of heavy-ion systems, supports the above conclusion and also brings out in an unambiguous way that the observed energy loss can be accounted primarily by the nucleon exchange mechanism alone without having to invoke any arbitrary parameter. For most of the systems including $\text{Xe} + \text{Bi}$ and $\text{Pb} + \text{U}$ where the target or the projectile nuclei have significant negative shell correlation energies, there is
no evidence for significant shell effects in the observed correlations between the energy loss and $\sigma_Z^2$. However, for Pb + Pb system, some deviations from the predictions of the model are observed which need further investigations.

In the above discussion, we have focussed on the expected correlation between energy loss and the variance of the fragment mass and charge distributions in heavy ion deep inelastic collisions. But there are other features of the heavy ion collisions which are also governed by this basic mechanism. The transfer of the orbital angular momentum to the fragment spins, and the misalignment of the transferred spins with respect to the reaction plane have also been discussed in the literature (Randrup 1983) on the basis of the nucleon exchange mechanism. In a recent study (Ramamurthy and Kapoor 1984) aimed at understanding the observed anomalous fragment angular distributions in heavy ion induced fusion-fission, where the relaxation of the dinuclear complex in the mass-asymmetric degree of freedom is reached during fusion, the analysis of the data brings out that such a complex relaxes in the K-degree of freedom on a time scale of several times $10^{-21}$ sec, and the fused composite nuclei with fission barriers comparable to temperature can fission from the non-equilibrium state of
unrelaxed K-distribution. Presence of this type of non-equilibrium fission is shown to be responsible for the apparently anomalous fragment angular distributions. The importance of the role of nucleon-exchange mechanism in the process of K-equilibration has been recognized, but a quantitative study is needed to arrive at more definite conclusions. There are several other aspects of nucleon exchange mechanism in fission and heavy-ion reactions such as the effects of cluster transfers which are also currently being explored.

In conclusion, the nucleon exchange mechanism operating between two nuclei in proximity first suggested for the fission process has now been extensively studied through the deep inelastic heavy ion collisions. The experimentally well-established correlation between the dissipated energy and the variance of the fragment charge or mass distribution in the deep inelastic heavy ion collisions, and the transport model description taking into account neutron-proton correlations in the exchange process, bring out in an unambiguous way the dominant role of the nucleon exchange process in the energy loss mechanism in heavy ion reactions. It is also recognized that several other observed features in heavy ion reactions are also governed by this underlying mechanism.

References


Gobbi A and Norenberg W 1980 *Heavy ion collisions* (ed.) R. Bock (Amsterdam, North Holland) vol. 2, p. 128

Kapoor S S 1982 *Curr. Sci.* 51 913


Ramamurthy V S 1971 Ph.D. thesis Bombay University


Ramanna R, 1964 *Phys. Lett.* 10 321


Randrup J 1977 *Nucl. Phys.* A327 490

Randrup J 1983 *Nucl. Phys.* A383 486


Weiden Muller H A 1980 *Progr. in Part. Nucl. Phys.* 3 49

Angular distribution in ternary fission

M M SHARMA and G K MEHTA
Department of Physics, Indian Institute of Technology, Kanpur 208 016, India

Abstract. The angular distribution of long-range alpha particles emitted in keV-neutron induced fission of $^{235}$U has been measured using a technique which employs only a particle telescope to derive the angular information. The neutron energy region investigated is 100 keV–1 MeV. The angular distribution of LRAs has been found to be peaked perpendicular to the neutron-direction with a substantial amount of anisotropy near 200 keV.

Keywords. Fission; LRA-fission; angular distribution.

PACS No. 25 85

1. Introduction

Light-charged particle (LCP) accompanied fission represents an interesting phenomenon in the study of the fission process. It is a potential source of information on the configuration of the fissioning nucleus at scission and the subsequent dynamics of the process till the last stage of rupture. The striking feature of the LCP-emission is that their angular distribution is strongly peaked perpendicular to the fission-fragment direction. This has led to the belief that they are born in the vicinity of the scission point from the neck of the compound nucleus. Several hypotheses on the emission mechanism of the LCPs viz the pre-scission evaporation model (Ramanna et al 1963), Carjan’s model (1975) of pre-scission emission, the post-scission emission model by Feather (1969), the sudden-snap model (Halpern 1965) and the statistical model (Fong 1970) are in vogue. However, it is not yet established whether the LCPs are emitted before or after the scission. The experiments on the angular distribution of the light-charged particles (LCPs) about the fission axis have yielded some information on the configuration of the nucleus at the scission point with the help of trajectory calculations. However, there is always an uncertainty in locating the scission point in trajectory calculations which reflects in the non-uniqueness of the information derived about the dynamical variables at scission. The angular distribution of the LCPs about the space-fixed axis (incident neutron direction) could provide valuable information on the emission mechanism of these LCPs. This aspect of the LCPs is not studied well and the existing experiments have yielded conflicting results. The present paper is an effort to understand the LCP-emission mechanism by studying the anisotropy of LCPs about the incident projectile direction.

The angular distribution of the binary fission fragments in neutron-induced fission near the fission threshold exhibits a forward peaking about the incident neutron direction. On the other hand, the trend of the fragment angular distribution in ternary fission (fission accompanied by a third particle (LCP)) is unclear. The behaviour of the LCP angular distribution should depend on the fact whether the LCP emission takes place...
before or after the scission. The angular distribution of the fragments and the lras (long-range alpha particles, the most predominant lcp) in 14 MeV neutron-induced fission of $^{235}$U was measured by Ramanna et al (1963). They observed that the lra angular distribution was peaked fore and aft along the beam direction, and the associated fragments (referred to as ternary fragments) were peaked at 90° to the beam. This behaviour of ternary fragments was opposite to the behaviour of fragments in binary fission. This result implied that the ternary fission process is distinctly different from the binary process notwithstanding that all the other characteristics of the fission process, such as kinetic energy distribution, mass distribution etc are known to be very similar in the two cases of the binary and ternary fission. However, in another measurement, 17.5 MeV proton-induced fission of $^{238}$U, Atneosen et al (1965) observed that the ternary fission fragments have nearly the same (0°-peaked) angular distribution as the binary fission fragments. The results of Ramanna et al (1963) and Atneosen et al (1965) are contradictory to each other. In both these experiments there is substantial contribution from second and third chance fissions, and the observed angular distribution arises from contributions due to a number of nuclides fissioning at different excitation energies. Nadkarni (1968) measured the angular distribution of ternary fission fragments and lras about the neutron-beam direction in 3 MeV neutron-induced fission of $^{235}$U. The ternary-fragment anisotropy was found to be opposite to that of binary fragments, in agreement with the results of Ramanna et al (1963). The results were interpreted by assuming that the lra emission takes place as a result of evaporation before the scission point. However, it was not understood as to why in this particular characteristic, the ternary fragments behave in a completely different way than the binary fragments whereas in all other characteristics they are similar. In view of the unclear picture of ternary-fission angular distribution as stated above, we have measured the angular distribution of lras in the neutron-induced fission of $^{235}$U for several energies in the region 100 keV to 1 MeV.

2. Experimental technique

The low probability of occurrence of the ternary fission, particularly in fast-neutron induced fission, makes it difficult to investigate the angular distribution of the lras using conventional techniques. We have developed a method (Sharma et al 1984) for determining the angular distribution using a $\Delta E - E$ particle telescope. This allows simultaneous identification of the particle and determination of its angular distribution. The energy loss in the $\Delta E$-detector is proportional to thickness $\Delta x$ of the detector for the normal incidence. However, if the particle is incident at an angle $\theta$ with the telescope axis, it suffers an energy loss proportional to the slant thickness traversed in the $\Delta E$-detector and therefore, $\Delta E$-signal carries an information on the angle $\theta$. Using the range-energy tables, the angular distribution is extracted, from the relation

$$\cos \theta = \Delta x / \Delta x',$$  

(1)

where $\Delta x$ and $\Delta x'$ refer to the normal thickness and the slant distance traversed by the particle in $\Delta E$-detector respectively.
3. Experimental details

The experiment used a 97% enriched $^{235}$U source of thickness 5 mg/cm$^2$ with an active area 4 cm$^2$ and a semi-conductor detector particle telescope consisting of 50 μm thick $\Delta E$-detector and a 500 μm thick $E$-detector. An aluminium foil was placed between the $^{235}$U source and the particle telescope to stop the natural alpha particles and the fission fragments from reaching the detectors. The space-fixed axis (neutron-beam direction) coincided with the telescope axis. The distance of the $^{235}$U source from the neutron producing target was such that the neutron divergence was $\approx 10^\circ$. The neutrons were produced using $^7$Li$(p, n)$ and $^3$H$(p, n)$ reactions with 2 MV Van de Graaff accelerator at Indian Institute of Technology, Kanpur. The data were recorded at the following neutron energies: thermal, $140 \pm 30$ keV, $170 \pm 25$ keV, $200 \pm 25$ keV, $400 \pm 200$ keV, $600 \pm 180$ keV and $1000 \pm 170$ keV. The large spread at higher energies was owing to the use of 6 Ci tritium target. The data were recorded event-by-event on a magnetic tape using a 3-parameter data-acquisition system and the analysis was done off-line on DEC-1090 computer.

4. Data analysis and results

The added advantage of our technique is that it allows the use of broad area source which helps in overcoming the count-rate problem. However, with broad source the contribution to the angular distribution arises from every point of the source, thereby causing a smearing in the measured angular distribution. The actual angular distribution is disentangled from the experimental one by Monte-Carlo simulation with the given geometrical constraints. Figure 1 shows the calculated (Monte-Carlo) and experimental angular distributions for thermal-neutron induced fission. The calculated distribution has been obtained for isotropic emission in thermal neutron-induced fission. The calculated distribution has been obtained for isotropic emission in thermal neutron-induced fission. The calculated and experimental distributions show excellent agreement. The curves can be fitted to straight lines for angles beyond 10$^\circ$. The points upto 10$^\circ$ have been excluded because of large errors due to $(\sin \theta)^{-1}$ factor near $\theta = 0^\circ$ while calculating $d\sigma/d\Omega$. The least-square fitted straight lines to the calculated and experimental curves give identical values of slopes. The slopes of the straight lines

![Figure 1. Calculated (Monte-Carlo) and experimental angular distributions of $^{133}$S emitted in thermal-neutron induced fission.](image-url)
depend upon the anisotropy of the angular distribution. Monte-Carlo calculations for various anisotropic distributions were carried out. The distributions were assumed to go as

\[ W(\theta) = 1 + \alpha \cos^2 \theta, \quad (2) \]

where \( \alpha \) represents the anisotropy, the positive value indicating a fore and aft peaking and a negative one showing perpendicular peaking. Figure 2 shows the variation of slopes with negative values of \( \alpha \), as all the experimental curves in the fast-neutron fission showed a decrease in the slope compared to thermal-neutron fission. This corresponds to a negative value of \( \alpha \) and the perpendicular peaking. For quantitative determination of anisotropy, figure 2 served as a calibration curve. The anisotropies thus determined are shown in table 1. Our anisotropy results lead to the conclusion that the \( \text{LRA} \) angular distribution is peaked perpendicular to the neutron-beam direction at all neutron energies from 140 keV to 1 MeV (figure 3). Assuming that the alpha-fragment angular

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2.png}
\caption{Slope of the least-squares fitted straight lines as a function of anisotropy (\( \alpha \)).}
\end{figure}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Energy of incident neutron (keV) & Anisotropy (\( \alpha \)) (\%) \\
\hline
140 \( \pm \) 30 & 85 \( \pm \) 28 \\
170 \( \pm \) 25 & 87 \( \pm \) 32 \\
200 \( \pm \) 25 & 94 \( \pm \) 31 \\
400 \( \pm \) 200 & 10 \( \pm \) 28 \\
600 \( \pm \) 180 & 25 \( \pm \) 19 \\
1000 \( \pm \) 170 & 50 \( \pm \) 27 \\
\hline
\end{tabular}
\caption{Anisotropy for several incident neutron energies.}
\end{table}
Angular distribution in ternary fission

Figure 3. Anisotropy of the \textit{lra}s at different neutron energies.

correlation does not undergo any change from thermal to fast fission, the ternary-fragment angular distribution in our experiment will be peaked forward and backward to neutron direction as does the binary-fragment angular distribution. We find that the \textit{lra} anisotropy is very high near 200 keV neutron energy. The error bars indicated in anisotropy in figure 3 are due to statistical origin only as the angular resolution due to our technique is only $\approx \pm 3^\circ$ and better above angle $15^\circ$.

5. Discussion

The angular distribution of the long-range alpha particles (\textit{lra}s) emitted in ternary fission is observed to be peaked perpendicular to the incident neutron direction. This implies that the ternary fission fragments are forward-peaked which is similar to the case in the binary fission. Thus it is concluded that the characteristics of the fission fragments are similar in both binary and ternary fission in all aspects. The similarity in angular distribution suggests that the $K$-distributions at the saddle point are the same whether a third particle accompanied the fission or not. Earlier measurements of Ramanna \textit{et al} (1963) and Nadkarni (1968) do not agree with our results. However, these angular-distribution measurements were for higher energy neutron-induced fission. They had proposed that the \textit{lra} emission takes place as evaporation before the scission point. Our results on ternary fission in the neutron energy range of 100 keV to 1 MeV lend credence to the assertion that the \textit{lra} emission takes place near the scission point. The models which can predict the angular distributions in conformity with our results are the sudden-snap model (Halpern 1965) and the statistical model (Fong 1970).

We find that the anisotropy is very high near 200 keV neutron energy and then reduces at higher energies. In the region around 200 keV, the interaction of $p$-wave
neutrons is predominant and the states $2^+, 3^+, 4^+$ and $5^+$ are accessible to the fissioning nucleus at the saddle point. The increased anisotropy seems to be due to these positive parity states at the transition point. At higher neutron energy, other partial waves also contribute significantly and thus one observes the angular distribution averaged over states of both parities. This may be the reason for decrease in anisotropy at higher energies.

Acknowledgements

We thank the technical staff of our Van de Graaff Laboratory for their assistance during the experiment. Discussions with Dr Ajit Sinha are appreciated. This work was partially supported by a research grant from the Department of Atomic Energy, Government of India.

References

Atneosen R A, Thomas T D and Garvey G T 1965 Phys. Rev. 139 B307
Feather N 1969 Proc. R. Soc. Edinburgh A68 229
Radiochemical studies on fission of actinides

M V RAMANIAH
Bhabha Atomic Research Centre, Trombay, Bombay 400085, India

Abstract. Since its discovery in 1939, nuclear fission has been extensively studied by various experimental as well as theoretical groups in several countries leading to an understanding of major aspects of this important and complex nuclear reaction. In Trombay, studies have been carried out in the last 25 years using both physical and radiochemical methods and significant contributions have been made towards a better understanding of this reaction. This paper presents highlights of radiochemical studies on fission of actinides, particularly mass, kinetic energy and charge distribution and fragment angular momentum. Results of these studies brought out the important role played by deformation energy surface, spherical and deformed nuclear shells and nucleon pairing.

Keywords. Nuclear fission; radiochemical methods; mass; charge; kinetic energy distributions; fragment angular momenta; nuclear shells and nucleon pairing.

PACS No. 25-85

1. Introduction

The high sensitivity and specificity of radiochemical methods led to the discovery of nuclear fission by Hahn and Strassmann in 1939. Since then, all aspects of this complex nuclear reaction have been studied extensively by radiochemical as well as physical methods. Soon after the discovery, Böhr and Wheeler (1939) put forward a theory of nuclear fission based on the liquid drop model (LDM), which was able to explain some broad features of the reaction such as energetics of fission process, fission probability and fissionability. The most striking feature of fission, namely, asymmetric mass distribution could not be explained by the LDM, according to which the mass distribution in low energy fission should be symmetric. This problem evoked great interest in several experimental and theoretical groups in the world. Within about a decade since the discovery of fission, a massive amount of experimental data was collected on mass, charge and kinetic energy distribution as well as other aspects. Considerable effort was directed towards understanding the variation of potential energy surface as a function of symmetric and asymmetric deformation for a wide range of fissionability parameters with the aim of providing an explanation for the observed mass asymmetry (Wilets 1964). Results of extensive experimental work covering mass, charge and kinetic energy distribution as well as prompt neutron and gamma emission brought out the importance of nuclear shells.

Several theoretical approaches were made incorporating shell effects, prominent among them being the statistical model (Fong 1956), and the nucleon exchange model (Ramanna 1964). As discussed by Kapoor and Ramamurthy in a paper to this volume, nucleon exchange model can explain one-, two-, and three-peaked mass distributions in terms of exchange of nucleons taking into account the influence of fragment nuclear shells. An important breakthrough came from the work of Myers and Swiatecki (1967)
and Strutinsky (1967). Incorporation of single particle effects over and above that of liquid drop potential by these groups resulted in a double-humped fission barrier in contrast to a single-humped barrier resulting from LDM consideration. Moller and Nilsson (1970) showed that in the vicinity of the second saddle point, asymmetric deformation of the fissioning nucleus results in the lowering of the fission barrier, which provided a qualitative explanation for the asymmetric mass distribution in terms of the double-humped fission barrier. The effect of washing out of nuclear shell effects with excitation energy on the observable quantities like the fragment angular distribution and fission probabilities in the case of nuclei with double-humped barrier was later worked out by Ramamurthy et al (1970). Mustafa et al (1973), based on the two-centre shell model, showed that the shell structure in the fragments influences the potential energy surface, explaining qualitatively the mass distribution of a wide range of fissioning nuclei. More recently, Wilkins et al (1976) and Prakash et al (1979) had attempted to explain various features of low energy fission as a result of recognising the importance of deformed nuclear shells in addition to spherical shells at scission in their model based on the quasi-statistical equilibrium near scission. A large volume of results of the theoretical and experimental studies of low-energy fission have been published in a number of excellent reviews and proceedings of International Conferences (Hyde 1966; Vandenbosch and Huizenga 1974; Proc. of the IAEA Symposia on Physics and Chemistry of Fission; 1965, 1969, 1973 and 1979).

At Trombay, work on nuclear fission was started by Dr Ramanna and it is because of his continued interest and strong support that basic research in this important field could be pursued and significant contributions made. Both physical and radiochemical methods have been employed in these investigations.

The areas of fission research to which a substantial contribution has been made by the Trombay group using physical techniques such as ionization chambers, scintillation detectors, semiconductor detectors and time of flight are the studies of fission fragment characteristics and secondary radiations emitted in fission. The early investigations concerned the mass, kinetic energy and angular distributions of fission fragments and the correlations among them (Kapoor et al 1965; Rekha Govil et al 1983), and the emission of neutrons (Ramanna et al 1961; Kapoor et al 1963), gamma rays (Kapoor and Ramanna 1964) and K-x-rays (Kapoor et al 1968, 1969, 1971; Kataria et al 1970) in coincidence with the fission fragments. In the last few years, a comprehensive programme to study the light-charged particles in fission has also been carried out (Ajitanand et al 1975; Choudhury et al 1976, 1980), the results of which have been reviewed elsewhere (Nadkarni 1982).

Radiochemical studies on the fission of actinides were initiated with the aim of understanding the systematics of low energy fission, particularly distribution of mass, charge, kinetic energy and angular momentum covering a wide range of elements. The results contributed to a better understanding of this interesting nuclear reaction, particularly the role played by both the spherical and deformed nuclear shells and the effect of nucleon pairing. In this article, we present a summary of the Trombay radiochemical work highlighting the important findings.

2. Mass distribution

At Trombay, mass distribution studies were carried out in the neutron-induced fission of a number of actinides ranging from $^{227}$Ac to $^{245}$Cm. These studies involved
Radiochemistry of fission of actinides

Radiochemical separations of fission products followed by their estimation using beta or gamma counting or direct counting of catcher foils on a high resolution germanium spectrometer. Several methods such as absolute method, relative method and comparison method (Jain and Ramaniah 1973) were employed in these studies. Some relevant results are presented in table 1. One interesting feature which emerged from the extensive data obtained at Trombay and from the literature data covering a wide range of fissioning nuclei from $^{227}$Ac to $^{259}$Md is the transition from three-peaked distribution in the Ac-Th-U region to two-peaked distribution in heavier actinides and a single-peaked mass distribution in the Fm-Md region. The transition from three-peaked to two-peaked distribution in Ac-Th region (Jensen and Fairhall 1958) was explained on the basis of two-mode fission hypothesis put forward by Turkevich and Niday (1951) according to which the three-peaked mass distribution is a result of superposition of asymmetric and symmetric mass distributions and transition to an asymmetric mass distribution occurs rather sharply in going from actinium to thorium.

Neutron-induced fission of $^{232}$Th (Iyer et al 1963) was studied at Trombay and a small symmetric peak was also observed besides the two large asymmetric peaks indicating that the extent of symmetric contribution may not strongly depend on the proton number of the fissioning nucleus. This was further confirmed by the mass distribution in the neutron fission of $^{232}$U (Manohar et al 1979) which showed a symmetric peak similar to that in the neutron fission of $^{232}$Th. The dependence of symmetric contribution on proton and neutron number of the fissioning nucleus was explained on the basis of the influence of variation of the difference between barrier heights for symmetric and asymmetric modes (Es-Ea) obtained from the work of Moller and Nix (1974) as a function of proton and neutron number as shown in figures 1a and 1b. It is seen that (Es-Ea) varies more sharply with the neutron number than the proton number of the fissioning nucleus and the contribution of the symmetric component increases as (Es-Ea) decreases. The observability of the symmetric peak in $^{232}$U could be understood as due to a decrease in (Es-Ea) with decrease in the neutron number from 143 to 141, which more than compensates for the increase in (Es-Ea) with increase in the proton number from 90 to 92.

Table 1. The mass distribution characteristics for different fissioning nuclei studied at Trombay

<table>
<thead>
<tr>
<th>Fissioning nucleus</th>
<th>Most probable mass</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Light</td>
<td>Heavy</td>
</tr>
<tr>
<td>$^{228}$Ac</td>
<td>88.0</td>
<td>138.0</td>
</tr>
<tr>
<td>$^{230}$Th</td>
<td>88.5</td>
<td>139.5</td>
</tr>
<tr>
<td>$^{233}$Th</td>
<td>93.0</td>
<td>137.5</td>
</tr>
<tr>
<td>$^{232}$Pa</td>
<td>94.2</td>
<td>139.8</td>
</tr>
<tr>
<td>$^{233}$U</td>
<td>94.1</td>
<td>138.9</td>
</tr>
<tr>
<td>$^{234}$U</td>
<td>95.0</td>
<td>139.0</td>
</tr>
<tr>
<td>$^{236}$U</td>
<td>95.5</td>
<td>139.5</td>
</tr>
<tr>
<td>$^{238}$Np</td>
<td>98.0</td>
<td>140.0</td>
</tr>
<tr>
<td>$^{240}$Pu</td>
<td>100.4</td>
<td>139.6</td>
</tr>
<tr>
<td>$^{242}$Pu</td>
<td>102.6</td>
<td>139.4</td>
</tr>
<tr>
<td>$^{244}$Cm</td>
<td>105.3</td>
<td>140.7</td>
</tr>
</tbody>
</table>

Figure 1a. Variation of \( (E_s - E_a) \) as a function of proton number of the fissioning nucleus (Manohar et al. 1979; values of \( E_s \) and \( E_a \) taken from Moller and Nix 1974).

Figure 1b. Variation of \( (E_s - E_a) \) as a function of neutron number of fissioning nucleus (Manohar et al. 1979; values of \( E_s \) and \( E_a \) taken from Moller and Nix 1974).
Figure 2. Variation of peak to valley ratio as a function of mass of fissioning nucleus in low energy fission (Manohar et al. 1979).

Figure 3. Mass yield curve in reactor neutron induced fission of $^{238}\text{U}$ extrapolated beyond $A = 70$ on lighter wing side and $A = 161$ on heavier wing side. Experimental points show the presence of new “shoulders” or “humps” in highly asymmetric region due to 28 proton shell effect (Rao et al. 1979).
This dependence of \((E_s - E_a)\) on proton and neutron number of the fissioning nucleus was also able to explain the variation of the peak-to-valley ratio, \(P/V\) (the ratio of the highest yield to the yield corresponding to the symmetric mass split), with mass number of the fissioning nucleus as shown in figure 2. It can be seen that \(^{236}\)U has the highest \(P/V\) ratio, corresponding to the maxima in \((E_s - E_a)\) around \(Z = 92\) and \(N = 146\). For fissioning nuclei heavier than \(^{236}\)U, \((E_s - E_a)\) decreases again, but the symmetric peak is not observed any more as it is obscured by the broadening of the mass distribution and the shift in the position of light wing towards heavier mass.

Another interesting finding at Trombay is the influence of 28 proton shell in enhancing the yields of fission products at highly asymmetric mass region 66–67 and the complementary masses as shown in figure 3 (Rao et al 1979). The measured yields in the range \(10^{-3} - 10^{-6}\%\) are about 1–3 orders higher than those expected from a smooth curve.

3. Kinetic energy distribution

A major part of energy released in low energy fission appears in the form of fragment kinetic energy. Studies on kinetic energy distribution provide information on the influence of fragment nuclear shells, fragment deformation and scission point configuration of the fissioning nucleus. Extensive studies have been carried out in this area using physical methods and time-of-flight measurements. The physical methods have limited sensitivity in the low fission yield region and in the case of target nuclei with high alpha specific activity. Due to their high sensitivity, radiochemical methods have been widely used particularly where physical methods have limitations. In the radiochemical methods, kinetic energies of fission products are calculated from the measured recoil ranges in a suitable stopping medium such as Al, air, etc using empirical range-energy equations. The kinetic energies of fission products are converted into those of fission fragments by appropriate corrections for neutron evaporation. This procedure gives accurate kinetic energy values when range-energy equations are calibrated using kinetic energy data obtained by physical methods on high yield fission products.

Kinetic energy distribution in low energy fission shows a maximum around mass 132, but not at the symmetric mass split as expected from the LDM. The difference between maximum kinetic energy and the kinetic energy corresponding to symmetric split known as kinetic energy deficit \((\text{kED})\), is explained on the basis of compact scission configuration of the fissioning nucleus for mass split corresponding to 132 (due to the presence of the doubly magic spherical shell with \(Z = 50\) and \(N = 82\)) resulting in a shorter distance \((D)\) between charge centres of the complementary fragments compared to that for a symmetric mass split.

Kinetic energy distribution in the neutron fission of \(^{232}\)Th, \(^{232}\)U, \(^{233}\)U, \(^{237}\)Np, \(^{239}\)Pu, \(^{241}\)Pu, \(^{241}\)Am and \(^{245}\)Cm was studied leading to estimates of \(\text{kED}\) and the average total kinetic energy \((\text{tKE})\) (Satya Prakash et al 1969, 1972a, b; Dange et al 1975; Ramaswami et al 1977). A linear correlation was found to exist between \(\text{kED}\) and the difference of fragment mass for symmetric split from the doubly magic number 132 as shown in figure 4. An extrapolation of this correlation indicates that \(\text{kED}\) for the fission of \(^{264}\)Fm tends towards zero and this trend was confirmed experimentally in the fission of \(^{258}\)Fm and \(^{259}\)Fm (Flynn et al 1978; Ragaini et al 1974).
The kinetic energy data from the large number of fissioning nuclei mentioned above was used to calculate the relative deformation and deformation parameters of various fission products to examine the dependence of deformation parameter on the fissionability parameter $X$. For this purpose the $D$ values for various mass splits were first calculated from the experimental values of kinetic energies. The $D$ values for a particular mass split was then apportioned between the complementary fragments on the basis of the proximity of their neutron and proton number to the nearest closed shell. Figure 5 shows a typical plot of deformations of fragments as a function of their masses. It was found that fragment deformation increases with increase in $X$ for
symmetric mass division, but does not change appreciably for asymmetric mass division. This difference could be explained on the basis of cylindrical shape of saddle point for nuclei with $X \geq 0.7$ which takes longer time to arrive at scission point with considerable stretching of the neck with increasing $X$. Events leading to symmetric splits seem to be mainly governed by liquid drop behaviour of the fissioning nucleus while asymmetric splits are influenced by fragment shells and hence remain insensitive to the variation of $X$. From the calculated fragment deformation, the total experimental kinetic energy and the calculated $Q$ values, the deformation energies of the fragments were calculated leading to estimates of the prompt neutron multiplicities as a function of fragment mass. Typical results for thermal neutron fission of $^{235}$U are shown in figure 6.

4. Charge distribution

Studies on charge distribution provide considerable insight into the influence of nuclear shells and nucleon pairing and dynamical aspects of descent of the fissioning nucleus. The isobaric and isotopic yield distributions are of Gaussian nature characterised by the most probable charge, $Z_p$ or mass $A_p$ and the corresponding width parameters, $\sigma_Z$ or $\sigma_A$.

Experimental investigations on charge distribution involve determination of the fractional cumulative yields ($FCY$) or fractional independent yields ($FIY$) of isobaric fission products or conversion of the independent yields of fission products to fractional isotopic yields making use of elemental yields. The physical methods involve on-line measurements of x-ray or $\gamma$-ray intensities in coincidence with kinetic energies of pairs of complementary fission products or use of mass separators, while
radiochemical methods are off-line and involve separation of fission products of interest and their assay by β/γ counting or direct gamma ray spectrometry. The necessity of the determination of the independent yields of fission products of short half-lives was a limitation of radiochemical methods in the past. However, the advent of fast and automated transport and radiochemical separation techniques during the last decade has attenuated this problem.

In our laboratories, isobaric charge distribution studies have been carried out in the mass region 100–105 and 130–140 in the neutron fission of $^{229}$Th, $^{239}$Pu, $^{241}$Pu and $^{245}$Cm (Rattan et al 1983; Datta et al 1980; Ramaswami et al 1982) and in the spontaneous fission of $^{252}$Cf (Manohar et al 1978; Srivastava et al 1984). In addition, investigations have also been carried out on the isotopic yields distribution of technetium in the neutron fission of $^{239}$Pu and spontaneous fission of $^{252}$Cf and of iodine isotopes in $^{252}$Cf(sf) and in 30 and 40 MeV alpha-induced fission of $^{232}$Th (Reddy et al 1984).

Spontaneously fissioning nuclei are ideal to understand the role of potential energy surface of fissioning nucleus and fragment nuclear structure on charge distribution. The fcy or fiy of a number of isobars obtained in our laboratory along with the data from literature (Wahl 1983) were analysed to obtain the values for $Z_p$ and $\sigma_z$ for mass chains 105, 131, 139, 140 and 141. It was found that the Gaussian distribution with $\sigma_z = 0.60 \pm 0.06$ represents most of the yields in $^{252}$Cf as shown in figure 7. It was

![Figure 7. A plot of fractional cumulative yields (fcy) as a function of $(Z-Z_p)$ in the spontaneous fission of $^{252}$Cf (Srivastava et al 1984; Srivastava 1984).](image-url)
further observed that in the mass region 130–135, $\sigma_z$ value is on the lower side (i.e. $\sigma_z = 0.50$ to 0.55) indicating sharper charge distribution in this mass region, apparently due to preferential formation of 82 shell fragments. An estimate of proton pairing effect ($\pm 12\%$) was obtained based on our data on elemental yield of iodine and literature data on other elements in the spontaneous fission of $^{252}\text{Cf}$.

From the data presented in figure 8 for $^{252}\text{Cf}$, it is seen that the deviation of $Z_p$ from $Z_{\text{UCD}}$ ($\Delta Z$) increases with mass asymmetry as expected from the minimum potential energy hypothesis. From a similar analysis of variation of $\Delta Z$ with mass asymmetry in the neutron fission of $^{233}\text{U}$, $^{235}\text{U}$, $^{239}\text{Pu}$ and $^{249}\text{Cf}$ (Wahl 1983; Mariolopoulos 1981), it was noted that the rate of increase of $\Delta Z$ with mass asymmetry increases with fissionability parameter as shown in table 2. With increasing fissionability, the saddle

![Image](image.png)

**Figure 8.** Variation of charge polarization with fragment mass in the spontaneous fission of $^{252}\text{Cf}$ (Srivastava et al 1984).

<table>
<thead>
<tr>
<th>Fissioning nucleus</th>
<th>$\Delta Z/\Delta M$</th>
<th>Fissionability parameter $X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{236}\text{U}$</td>
<td>$0.010 \pm 0.005$</td>
<td>0.772</td>
</tr>
<tr>
<td>$^{234}\text{U}$</td>
<td>$0.015 \pm 0.005$</td>
<td>0.774</td>
</tr>
<tr>
<td>$^{240}\text{Pu}$</td>
<td>$0.015 \pm 0.005$</td>
<td>0.790</td>
</tr>
<tr>
<td>$^{250}\text{Cf}$</td>
<td>$0.025 \pm 0.002$</td>
<td>0.824</td>
</tr>
<tr>
<td>$^{252}\text{Cf}$</td>
<td>$0.036 \pm 0.006$</td>
<td>0.821</td>
</tr>
</tbody>
</table>

* $X = \frac{Z^2 A}{50.88 \left[ 1 - 1.7826 \left( \frac{A - 2Z}{Z} \right)^2 \right]}$

where $A$ and $Z$ are the mass and charge of the fissioning nucleus.
point shapes becomes more compact (as pointed out in §3) and the nuclear viscosity increases making the time of descent longer, permitting higher charge polarization (ΔZ) to achieve preferred configurations.

The elemental yields for technetium (Z = 43) in the neutron fission of $^{239}$Pu and spontaneous fission of $^{252}$Cf were deduced from the isotopic yield distribution data. From a comparison of these data with the data in the neutron fission of $^{235}$U, it is seen that technetium yield increases in the order of fissioning nuclei $^{236}$U < $^{240}$Pu < $^{252}$Cf as shown in table 3. The same table also shows that while the Z of complementary fragment recedes from magic shell closure at Z = 50 from $^{236}$U to $^{252}$Cf, the neutron number of complementary fragment approaches spherical neutron shell at N = 82 in the fissioning nuclei $^{240}$Pu and the deformed neutron shell at N = 88 in $^{252}$Cf. It is also seen that the neutron-to-proton ratio of the complementary fragment lies close to that of the fissioning nucleus as one approaches $^{252}$Cf indicating that shell effect is most strongly manifested when it is in phase with liquid drop energetics.

The investigation of isotopic yields of iodine (Z = 53) in 30-40 MeV a-induced fission of $^{232}$Th throws some light on the extent of mass/charge equilibration in fission. Table 4 shows the data on isotopic yield distribution parameter, $A_p$ and $\sigma_A$, as obtained in $^{232}$Th (α_{30-40 MeV}) and the same based on literature data in $^{235}$U (n_{th,f}). With increase in excitation energy the $A_p$ value approaches the value expected from unchanged charge distribution hypothesis. The increase of $\sigma_A$ with excitation energy was interpreted to be due to influence of temperature on mass/charge equilibration, after taking into account multichance fission, indicating that with increase of excitation energy the time of descent is shorter permitting less time for acquiring the preferred configuration as seen at lower excitation energies.

### Table 3. Most probable neutron ($N_p$) values for charge splits involving technetium and its complementary products in $^{235}$U(n_{th,f}), $^{239}$Pu(n_{th,f}) and $^{252}$Cf(sf).

<table>
<thead>
<tr>
<th>Fissioning nucleus</th>
<th>Element (Z)</th>
<th>$A_p' \pm \Delta A_p$</th>
<th>$N_p \pm \Delta N_p$</th>
<th>$Y_p$ percent</th>
<th>$Q^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{236}$U</td>
<td>Technetium (43)</td>
<td>109.14 ± 0.14</td>
<td>66.14 ± 0.14</td>
<td>0.092 ± 0.024</td>
<td>1.017</td>
</tr>
<tr>
<td></td>
<td>Indium (49)</td>
<td>126.86 ± 0.14</td>
<td>77.86 ± 0.14</td>
<td>1.022</td>
<td></td>
</tr>
<tr>
<td>$^{240}$Pu</td>
<td>Technetium (43)</td>
<td>108.35 ± 0.45</td>
<td>65.35 ± 0.45</td>
<td>8.00 ± 1.32</td>
<td>1.017</td>
</tr>
<tr>
<td></td>
<td>Antimony (51)</td>
<td>131.65 ± 0.45</td>
<td>80.65 ± 0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^{252}$Cf</td>
<td>Technetium (43)</td>
<td>109.10 ± 0.50</td>
<td>66.10 ± 0.50</td>
<td>16.90 ± 2.09</td>
<td>1.017</td>
</tr>
<tr>
<td></td>
<td>Cesium (55)</td>
<td>142.90 ± 0.50</td>
<td>87.90 ± 0.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $Q^* = (N_p/Z)_{fragment}/(N/Z)_{fissioning nucleus}$.

### Table 4. Iodine isotopic yield distribution parameters.

<table>
<thead>
<tr>
<th>Fissioning system</th>
<th>Excitation energy (MeV)</th>
<th>$A_p$ value</th>
<th>$\sigma_A$ expt.</th>
<th>$\sigma_A$ Calc. (in $^{236}$U*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{233}$U + n_{th}</td>
<td>6.54</td>
<td>135.9</td>
<td>1.01 ± 0.12</td>
<td>1.30</td>
</tr>
<tr>
<td>$^{233}$Th + α_{30 MeV}</td>
<td>23.4</td>
<td>135.09</td>
<td>1.60 ± 0.16</td>
<td>1.51</td>
</tr>
<tr>
<td>$^{233}$Th + α_{40 MeV}</td>
<td>33.2</td>
<td>134.8</td>
<td>2.37 ± 0.24</td>
<td>1.71</td>
</tr>
</tbody>
</table>
5. Angular momentum

Measurement of angular momentum of fission fragments provides information on scission configuration and the dynamics of motion past the second saddle as fragments acquire their angular momenta due to mutual coulombic torque for axially asymmetric scission configuration and thus depend on fragment deformation. Among the various methods for the estimation of fragment angular momenta, physical methods have certain limitations, e.g. measurements of prompt gamma angular anisotropy provide only the gross estimate of average angular momentum and measurements on prompt gamma multiplicity and energy are dependent on the choice of various parameters. Radiochemical method is the most suitable as it provides fragment angular momenta as a function of their mass and charge and has much less number of adjustable parameters. The radiochemical method involves determination of the independent yield ratios of the fission product isomers followed by statistical model based analysis (Vandenbosch and Huizenga 1960) to correct for fragment de-excitation to arrive at fragment angular momenta.

At Trombay angular momenta of fragments corresponding to the fission products isomeric pairs $^{95}$Nb and $^{132}$I in the neutron fission of $^{233}$U, $^{131,133}$Te in neutron fission of $^{241}$Pu and of $^{111}$Pd and $^{117}$Cd, $^{131,133}$Te and $^{134}$I in spontaneous fission of $^{252}$Cf have been estimated (Datta et al 1982; Guin et al 1983; Data 1984). Studies have also been carried out for $^{131}$Te and $^{133}$Te in the 30 to 40 MeV alpha induced fission of $^{232}$Th (Datta et al 1983).

Results of investigation in the neutron fission of $^{233}$U (as given in figure 9) showed

![Figure 9. Plot of fragment angular momentum vs fragment atomic number in $^{233}$U(n,f) (Datta et al 1982).](image-url)
that angular momentum of fragments around doubly magic shell region 132 (Z = 50, N = 82) decreases as corresponding atomic number deviates from 50 proton spherical shell contrary to expectation. This apparent anomaly is explained on the basis of the presence of the deformed shell (Z = 38) in the complementary fragment that tends to reduce the deformation, thus lowering the angular momentum of the corresponding heavy fragment as atomic number of the latter approaches Z = 54 from Z = 50.

In the spontaneous fission of $^{252}$Cf, it was observed that there exists an inverse correlation between fragment angular momentum and elemental yield with odd Z fragments having high angular momentum and lower yield compared to neighbouring even Z elements as shown in figure 10. This implied that for fragments with higher angular momentum, larger amount of energy remains tied up as rotational energy reducing the intrinsic excitation energy and in turn their yield according to the statistical model. Further, for odd Z elements high angular momenta indicates higher scission point deformation compared to their even Z neighbours in the magic shell region arising due to polarization of even-even core of protons by the odd-proton since in the spherical shell region of $N = 82$ nuclear deformation is more strongly affected by protons than neutrons as has also been concluded from recent theoretical calculations (Madsen and Brown 1984).

The observations on fragment angular momenta in the $^{252}$Cf(sf) show that the fragment angular momenta are correlated to the scission point deformation dependent, in turn, upon fragment nuclear structure. As a consequence one would expect a correlation between fragment angular momentum and prompt neutron number. However, lack of any such correlation as observed has been explained to be due to the

![Figure 10. Correlation of elemental yield and angular momentum in $^{252}$Cf (Datta et al to be published).](image)
fact that the prompt neutron number depends on total excitation energy of fragment after shape relaxation.

Fragment deformation at scission has been deduced based on simple theoretical calculation in terms of bending mode oscillation of the fissioning nucleus and statistical model (figure 11). The Bohr-Mottelson deformation parameters ($\beta$) for various fragments in the spontaneous fission of $^{252}\text{Cf}$ could be obtained. The values for $^{114}\text{Pd}$ ($N = 66$) and $^{142}\text{Xe}$ ($N = 88$) are 0.63 and 0.72 respectively which are theoretically

![Diagram of fission process](image)

**Figure 11.** Schematic of origin of fragment angular momentum due to precision bending mode oscillations.

<table>
<thead>
<tr>
<th>Fissioning system</th>
<th>Excitation energy ($E^*$) (MeV)</th>
<th>Initial angular momentum $\langle 12 \rangle 1/2 \hbar$</th>
<th>Fragment angular momentum (Bf) $^{131}\text{Te}$</th>
<th>Fragment angular momentum (Bf) $^{133}\text{Te}$</th>
<th>Calculated Bf value for $^{133}\text{Te}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{235}\text{U}(n_{th},f)$</td>
<td>6.54</td>
<td>3-4</td>
<td>$6.0 \pm 1.5^a$</td>
<td>$5.9 \pm 1.5^a$</td>
<td>—</td>
</tr>
<tr>
<td>$^{232}\text{Th}(\alpha_{30\text{MeV}},f)$</td>
<td>23.36</td>
<td>10-2</td>
<td>$7.5 \pm 0.7^b$</td>
<td>$6.5 \pm 0.4^b$</td>
<td>7.03</td>
</tr>
<tr>
<td>$^{232}\text{Th}(\alpha_{40\text{MeV}},f)$</td>
<td>33.20</td>
<td>14-4</td>
<td>$5.7 \pm 0.6$</td>
<td>$6.9 \pm 0.4$</td>
<td>$7.81$</td>
</tr>
</tbody>
</table>

$^a$ Sarantites et al 1965.  
$^b$ Imanishi et al 1976.
predicted deformation parameters for occurrence of deformed shells at these neutron numbers, in low energy fission. Investigations on 30 to 40 MeV alpha induced fission of 232Th indicated that the angular momenta of fragments in the magic shell region remains unchanged with excitation energy as shown in table 5. These observations have been interpreted in terms of interplay between tendency of increase in angular momenta at higher excitation energy as against stiffness of compound nucleus at higher angular momentum towards bending oscillation and shorter time of descent. In order to analyse the influence of the fissioning nucleus excitation energy and angular momentum, calculations were carried out for evaluation of fragment angular momenta. This calculation was based on retainment of angular velocity of Fermi gas type fissioning nucleus at saddle point and population of higher rotational states at higher excitation energy taking into account the effect of multichance fission. The agreements between calculated and observed angular momenta indicates that at higher energy fragment angular momentum is primarily dependent on the properties of fissioning nucleus at the saddle point.

6. Conclusion

Radiochemical studies of fission carried out at Trombay during the last 25 years have contributed to a better understanding of this highly interesting and complex nuclear reaction. Extensive investigations on mass distribution have shown that the one-, two-, and three-peaked mass distribution in low energy fission arise due to the relative contributions of the symmetric and asymmetric modes of fission which, in turn, depend on the difference in the barrier heights for these modes. It has been further shown that this difference strongly depends on the neutron number than the proton number of the fissioning nucleus and further, the variation of this difference with mass number of the fissioning nucleus explains the observed variation in the peak-to-valley ratio with the fissioning nucleus. Enhanced fission yields in the highly asymmetric mass could be understood as due to the influence of the 28 p-shell. Results of kinetic energy distribution demonstrated the influence of the doubly magic shell at mass 132 on the total kinetic energy and kinetic energy distribution. The fragment deformation calculated from the kinetic energy data has clearly shown that the symmetric mode is largely governed by the liquid drop behaviour. Results of studies on charge distribution and fragment angular momentum brought out the influence of 66 and 88 neutron deformed shells and further showed their greater effectiveness when in phase with liquid drop energetics. It was also shown that the extent of variation of charge polarisation with mass asymmetry depends on the fissionability parameter due to the influence of the saddle point shape and dynamics of descent. Angular momentum studies provided estimates of fragment deformation and showed that the odd Z fragments have high deformation due to the polarisation of even Z core by the odd proton. Results of studies on medium energy fission have shown the influence of the saddle point properties on charge distribution and fragment angular momentum due to higher excitation energy and shorter time of descent. Thus, our studies have contributed to a better understanding of the nuclear fission process particularly concerning the late stage of fission process and have brought out the importance of the deformation energy surface, spherical and deformed nuclear shells, nucleon pairing and the dynamics of descent past the second saddle point.
Acknowledgement

The author is grateful to the large number of his colleagues for their valuable help, suggestions and comments during the preparation of the paper. Particular mention should be made of Drs S S Kapoor, S B Manohar, P R Natarajan, V S Ramamurthy, Satya Prakash and Tarun Datta.

References

Bohr N and Wheeler J A 1939 Phys. Rev. 56 426
Datta T 1984 Fission studies on actinides, Ph.D. Thesis Bombay Univ.
Fong P 1956 Phys. Rev. 102 434
Hahn O and Strassmann F 1939 Naturwissenschaften 27 11
Hyde E K 1966 The nuclear properties of heavy elements (New Jersey Prentice Hall), Vol. 3
Jain H C and Ramaniah M V 1973 Radiochim. Acta 19 90
Myers W D and Swiatecki W J 1967 Ark. Fys. 36 343
Radiochemistry of fission of actinides

Ramanna R 1964 Phys. Lett. 10 321
Strutinsky V M 1967 Nucl. Phys. A95 420
Turkevich J A and Niday J B 1951 Phys. Rev. 84 52
Vandenbosch R and Huizenga J R 1960 Phys. Rev. 120 1305
Wahl A C 1983 (private communication)
Wilets L 1964 Theories of nuclear fission (Oxford: Clarendon Press)
Wilkins B D, Steinberg E P and Chasman R R 1976 Phys. Rev. C14 1832
Wing J and Fong P 1964 Phys. Rev. 136 B923
Studies of light-charged particle emission in fission at Trombay

S S KAPOOR and D M NADKARNI

Bhabha Atomic Research Centre, Trombay, Bombay 400085, India

Abstract. Studies of prompt radiations emitted in fission were started at Trombay in the late 1950's by Dr R Ramanna and over the years extensive investigations on the emission of prompt neutrons, gamma ray and K x-rays in fission were carried out with neutron beams from APSARA and CIRUS reactors. In the early 1960's studies on the emission of light-charged particles in fission, which is a rare mode of fission, were also started. This paper reviews some of the recent studies on the emission of light-charged particles (LCP) in fission which were carried out with a view to investigate the mechanism of LCP emission, the scission configuration and the dynamics of the last stages of the fission process.

Keywords. Nuclear fission; light-charged particle emission; emission probability; energy-angular-distribution.

PACS No. 25.85

1. Introduction

Nuclear fission process, which involves collective motion, large scale deformation and subsequent division of the fissioning nucleus into two fragments has provided over the years a convenient way to unfold macroscopic nuclear behaviour and to study nuclear dynamics. Only in recent years with the availability of energetic heavy ion beams the studies of nucleus-nucleus collision have provided a complementary and perhaps a more powerful way to investigate the domains of nuclear macrophysics which could earlier be probed only through the study of fission process. In the fission process once the nucleus is given sufficient energy to reach the top of the fission barrier (saddle point) it continues to elongate until it splits into two fragments at what we call “scission point”. The deformed fragment nuclei formed at scission convert almost instantly their deformation energies into excitation energy via damped vibrations around their stable shapes. Simultaneously, the resulting excited fragments undergo acceleration under the mutual Coulomb repulsion and acquire their full kinetic energies in a very short time ($\leq 10^{-20}$ sec) after scission. During the de-excitation of the fragments most ($\sim 90\%$) of the prompt neutrons are emitted after the fragments have reached their full kinetic energies. The excitation energy below the neutron binding energy is given off as prompt gamma rays via gamma transitions cascading down from the excited states to the ground state of the emitting fragment. A small fraction (about 6\%) of these transitions undergo internal conversion resulting in the emission of conversion electrons and the characteristic x-rays of the fragments. Occasionally, along with the two heavier fragment nuclei the fission process also results in the emission of a light-charged particle (LCP) and this process is called “LCP-accompanied fission” or simply as “ternary fission”. As the bulk of these light particles are long range alpha particles (LRA) this process is also often referred to as “LRA-accompanied fission” particularly when
measured without particle-identification of the LCP. As this process has a very low probability of occurrence of about $2 \times 10^{-3}$ as compared to the normal binary fission, detailed investigations of this process have been possible only for thermally fissile nuclei with the use of high flux reactors or for the spontaneous fission of $^{252}$Cf.

Investigations on the nuclear fission process were initiated at Trombay by Dr R Ramanna in the late 1950's and over the years considerable work has been carried out at Trombay with neutron beams from APSARA and CIRUS reactors on the various aspects of the fission process using physical and radiochemical techniques. The work using radiochemical techniques has been reviewed in another article by Dr Ramaniah in this volume. Among the earliest investigations in fission using physical techniques at Trombay were the measurements of the energy spectrum of the neutrons and neutron-fragment angular correlations in thermal neutron induced fission of $^{235}$U (Ramanna et al 1961; Kapoor et al 1963). The neutron-fragment angular correlations showed that a small fraction of neutrons ($\sim 10\%$ of total) does not originate from the de-excitation of the moving fragments and is emitted isotropically in the laboratory system with an evaporation-like energy spectrum. In this work, emission spectrum of the neutrons emitted from fully accelerated fragments was shown to be a superposition of various evaporation spectra with a distribution of nuclear temperatures of the form $P(T) = 2T/T_m$ with a maximum temperature $T_m$. This type of distribution has formed the basis of several recent theoretical calculations of the fission neutron spectra (Madland and Nix 1982). In a subsequent study the prompt gamma-fragment angular correlations were measured which showed an anisotropic gamma ray emission (anisotropy $\sim 10$–$15\%$) suggesting the occurrence of large angular momenta of fission fragments brought about possibly by the non-axial splitting of the neck-region of the fissioning nucleus at scission (Kapoor and Ramanna 1964). Several studies on the emission of K x-rays from fission fragments in spontaneous fission of $^{252}$Cf and thermal neutron induced fission of $^{235}$U were also later carried out. Among these were the measurements, for the first time, of the multiplicities of K x-rays emitted by the fission fragments of specified atomic numbers and the first and second moments of the x-ray emission distribution function (Kapoor et al 1971a,b). The average yields and emission times of K x-rays emitted from individual fission fragments were determined and from these the onset of a new region of nuclear deformation corresponding to neutron number $N > 88$ was established. It was also seen that there exists a significantly large probability of cascade emission of K x-rays in several cases.

Among other areas of fission physics research undertaken at Trombay were studies of fragment angular distributions (Nadkarni et al 1968b), correlation between fragment anisotropy and asymmetry in fast neutron-induced fission (Kapoor et al 1965) and correlation between fragment mass, charge and kinetic energy in thermal neutron fission (Govil et al 1983). We do not intend to discuss these investigations of binary fission process here but rather focus on some of the studies of the LCP-accompanied fission undertaken at Trombay.

The emission of LCP in fission has several characteristic features: (i) they are emitted with a very low probability ($P_{LCP}$) which is one in about 300 to 500 binary fissions, (ii) a majority of these LCP are isotopes of H and He, with almost $90\%$ of them being long range alpha particles (LRA), (iii) they are emitted with a broad energy spectrum with a near-Gaussian shape, which ranges from 5 to 30 MeV with a peak around 15 MeV and FWHM of about 10 MeV and (iv) they are emitted nearly perpendicular to the direction of fission fragment motion, more precisely at an average angle $\theta_{al}$ of about $83^\circ$ with...
respect to the direction of motion of the light fragment. It has been inferred from these characteristics that they are emitted close to the scission stage from the neck region between the two nascent fission fragments. In order to deduce information about the scission configuration as well as the dynamics of the fission process at the instant of their emission during the last stages of scission, more detailed information about the correlations among the various measured parameters is required. In the early 1960's Ramanna and coworkers started investigations on the emission of LCP in fission and the first study in this series was measurements of the angular distribution of the LRA and ternary fission fragments in 14 MeV neutron induced fission of $^{238}$U (Ramanna et al 1963). Numerous other studies of the LCP-accompanied fission followed in the subsequent years with the aim of investigating not only the mechanism of emission of these particles but also the scission configuration and the nuclear dynamics of the last stages of the fission process, using these LCP as probes. In what follows, some of the recent investigations on the LCP-accompanied fission carried out at Trombay are briefly reviewed.

2. Studies of LCP emission in $^{235}$U ($\text{n}_{\text{th}}, f$)

With a view to study the effect of the expected sharp variation of the Coulomb field around the inter-fragment (neck) region on the energy of the LRA, the energy spectra of LRA emitted at several angles with respect to the fragment direction were measured (Nadkarni et al 1968, 1972). From the energy spectra observed at these angles the width represented by $\sigma_\theta$ of the Gaussian-shaped LRA-fragment angular correlation function was deduced for LRA of different energies. These values of $\sigma_\theta$ vs LRA energy $E_\alpha$ are shown in figure 1 where a shallow minimum of $\sigma_\theta$ at $E_\alpha \sim 17$ MeV is observed. It was noticed that the yield of LRA does not decrease to a very low value at such small angles as 27° or 11° as would be expected for a true Gaussian distribution. Most of the LRA yield at these

![Figure 1. The variation of $\sigma_\theta$ of the Gaussian angular correlation function with alpha particle energy (Nadkarni et al 1972).](image-url)
small angles, belonging to a non-Gaussian component, shows a significantly higher most probable energy (22–24 MeV) (Nadkarni et al 1972). This non-Gaussian LRA component, which has been called polar LRA (Piasecki et al 1970) was attributed to evaporation from moving fission fragments which then explains the observed most probable energies at these small angles. Estimates of $\Gamma_a/\Gamma_n$ based on the statistical theory showed that at least some of the ‘polar’ LRA may be accounted for as evaporation from moving fission fragments (Nadkarni et al 1972).

Considerable insight into the mechanism of LCP emission in fission as well as of the fission process itself can be got from a knowledge of the kinetic energy $E_{Ko}$ of the fragments at the moment of liberation of the light-charged particles. Considering that LCP are liberated near the scission configuration, the quantity $E_{Ko}$ also corresponds to the prescission kinetic energy in the normal binary fission process. The magnitude of $E_{Ko}$ in the fission process is expected to be dependent on the degree of coupling of the collective and single particle degrees of freedom, and is therefore related to the magnitude of nuclear viscosity operating during the descent of the nucleus from saddle to scission. A direct determination of the magnitude of $E_{Ko}$ acquired by the fragments in a time of the order of few times $10^{-21}$ sec is not possible. But it has been realized for some time that the LCP can serve as probes to gauge this quantity through measurements of the various correlations in LCP accompanied fission (Frenkel 1967; Mehta et al 1973).

In the LRA-fission process several variables or parameters are involved such as the fission fragment mass ($M_f$), total fragment kinetic energy ($E_K$), LRA energy ($E_a$) and LRA-fragment angle $\theta_{sf}$. Although in earlier studies (Fraenkel 1967; Mehta et al 1973) correlation of $E_a$ and $E_K$ has been measured and analysed using trajectory calculations, a definite result on the magnitude of $E_{Ko}$ has been difficult to deduce from a limited set of measured correlations. As with more detailed data on differential correlation among all possible LRA-fission parameters it becomes possible to obtain more accurately the scission parameters and hence $E_{Ko}$, detailed multiparameter correlation measurements between the parameters $M_f$, $E_K$, $E_a$ and $\theta_{sf}$ in thermal neutron induced fission of $^{235}$U were carried out at Trombay with CIRUS reactor (Choudhury et al 1980). A back-to-back gridded ionization chamber (figure 2) was used to measure kinetic energies of the pair fragments and the angle $\theta_{sf}$ was measured by an electronic method based on the measurement of coincident pulse heights of the grid and the collector pulses (Choudhury et al 1979). Pulse heights of the two collectors, the two grids and of the LRA detector placed along the electric field direction of the chamber were recorded event-by-event on a magnetic tape by means of a 6-parameter data acquisition system. The data were analysed to obtain $M_f$, $E_K$, $E_a$ and $\theta_{sf}$ for each event and a probability matrix $P(M_f, E_K, E_a, E_r, \theta_{sf})$ was constructed, where total kinetic energy released is $E_r = E_K + E_a$, and $\theta_{sf}$ is the angle between LRA and the light fission fragment. From this matrix, the correlation between any two parameters for specific windows on the remaining parameters were obtained. Various correlations of interest were deduced and analysed in this work and we show here only a few selected results. Figure 3 shows the ternary fission fragment mass distribution, where the correction due to alpha particle recoil has been incorporated, together with the binary fission fragment mass distribution. The light fragment mass peak is observed to be shifted to the lower side whereas the heavy fragment peak is not shifted appreciably. This has been interpreted in terms of the
dependence of the alpha particle emission probability $P_a$ on fragment mass (Choudhury et al 1980). Figure 4 shows the variation of the most probable value of the total kinetic energy $E_T$ and $\sigma_{E_T}$ with heavy fragment mass $M_H$ in LRA and binary fission, where a significantly larger $\sigma_{E_T}$ is observed in LRA accompanied fission for $M_H \sim 130$ (near closed shell region) whereas nearly equal $\sigma_{E_T}$ is observed for $M_H \gtrsim 140$ (away from closed shell). These results were interpreted in terms of differences in the dependence of $P_a$ on the interfragment separation ($d$) for these two mass regions. An important correlation is the dependence of the width $\sigma_{\theta_{aL}}$ of the LRA-light fragment
angular correlation on the total fragment kinetic energy $E_K$ (figure 5). If the variation of $\sigma_{\theta_{SL}}$ with $E_K$ is examined under the two assumptions that the variation of the final $E_K$ can be due either to (i) the variation of the prescission kinetic energy $E_{Ko}$ or (ii) the variation in the Coulomb energy, the case (i) leads to a larger $\sigma_{\theta_{SL}}$ with increasing $E_K$ whereas the case (ii) leads to a smaller $\sigma_{\theta_{SL}}$ with increasing $E_K$. The observed $\sigma_{\theta_{SL}}$ vs $E_K$ favours the case (ii) implying that $E_{Ko}$ does not increase with $E_K$ which is expected only if $E_{Ko}$ has a small magnitude. Trajectory calculations based on three-point charge approximation have been done to fit these data to arrive at an initial set of scission parameters and their widths. From the above studies it has been concluded that the fission fragments have a very small prescission kinetic energy and that the alpha particle is liberated close to either of the two fragments and slightly off the axis joining the two fragment centres. This result has a direct relevance to the nature of fission dynamics and points to a strong coupling between the collective and internal degrees of freedom suggesting applicability of the statistical models during the saddle to scission.

In thermal neutron induced fission of $^{235}$U a still more rare mode of fission in which two light charged particles are emitted simultaneously together with the two heavier fragments was first observed at Trombay (Kapoor et al 1972). The probability of such events was estimated to be about one per million binary events and the gross features of LCP energy spectrum observed in this case were found to be similar to those in ternary fission. Later in spontaneous fission of $^{252}$Cf these LCP were identified as $^1$H, $^3$H and $^4$He and their relative yields were found to be similar to that in ternary fission. Although their energies were not strongly correlated their angles of emission were found to be well correlated (Kataria et al 1973). Trajectory calculations were done using different initial conditions and the one which assumed that the two LCP are emitted in the neck region statistically independent of each other at different times near scission gave a better fit to the experimental data.
3. Studies of \textit{LRA} emission in fast neutron induced fission

3.1 \textit{Emission probability of LRA and LCP in fission}

As the mechanism of LCP emission in fission is not well understood and very little was known about the factors which decisively influence the emission probability $P_{\text{LCP}}$, one of the aspects of LCP emission investigated in detail was the dependence of $P_{\text{LCP}}$ on the excitation energy of the fissioning nucleus, in $^{235}\text{U}(n,f)$. In this case the region of incident neutron energy of particular interest is thermal to about 5 MeV, because at much higher energies contributions from second and higher chance fissions complicate the interpretation of the results. The results of measurements at thermal, 2,3 and 4 MeV neutrons showed no appreciable change in the $P_a$ with neutron energy $E_n$ implying that if $P_a$ is dependent on the scission configuration or excitation energy at scission, these latter quantities do not change significantly with the initial excitation energy of the fissioning nucleus (Nadkarni and Kapoor 1970). In a later work the yield of different LCP, their energy spectra were measured in $^{235}\text{U}(n,f)$ for different values of $E_n$ using a $\Delta E-E$ detector telescope to identify the LCP (Sharma \textit{et al} 1981) using the 2 MV Van de Graaff machine at IIT Kanpur. In this study increase $\sim 20\%$ in $P_a$ for $E_n \sim 200$ keV was observed compared to that in thermal neutron fission in good agreement with an earlier measurement (Krishnarajulu \textit{et al} 1977). However, the yield of tritons increased with $E_n$ such that at $E_n \sim 550$ keV, it was about 3 times higher than the thermal neutron value. A still higher increase in the yield of protons was observed. These observations of
dependence of $P_{\text{LCP}}$ with $E_n$ have not yet found satisfactory explanation and need further investigations. Measurements were also carried out to study the dependence of the yield of polar and equatorial protons and alpha particles on incident neutron energy ($E_n \sim 600$ keV) where a similarity in the dependence of emission probability on the excitation energy of the fissioning nucleus was observed both for the polar and equatorial LCP (Sinha et al 1980, 1982, Sharma et al 1981).

3.2 Angular distribution of $\text{LRA}$ and ternary fission fragments with respect to the incident neutron direction

The angular distribution of $\text{LRA}$ and ternary fission fragments with respect to the incident neutrons was first measured by Ramanna et al in 14 MeV neutron fission of $^{238}$U. They observed a forward-backward peaked angular distribution of $\text{LRA}$ and a $90^\circ$-peaked distribution of ternary fission fragments. These results suggested that the $\text{LRA}$ emission is analogous to the evaporation of particles from a deformed and excited fissioning nucleus prior to scission. The observed results also implied that $\text{LRA}$ emission is dependent on the $K$-quantum number, $\text{LRA}$ emission being less probable for $K = 0$ than for the case where $K$ is large. In further investigations the anisotropy of the angular distribution of $\text{LRA}$ with respect to the incident beam was measured in 3 MeV neutron-induced fission of $^{235}$U (Hattangadi et al 1965) where the anisotropy was found to be in agreement with the value calculated on the basis of the statistical theory of evaporated particles. The observed $\text{LRA}$ anisotropy was compared with the anisotropy of the ternary fission fragments measured in 3 MeV neutron fission of $^{235}$U and it was found that the magnitude of the observed ternary fragment anisotropy is equal to that expected on the basis of the fact that $\text{LRA}$ are emitted prior to scission at right angles to the fission fragments (Nadkarni 1968). The question of $K$-dependence of the $\text{LCP}$ emission probability is also of current interest in the studies of $\text{LCP}$ emission in heavy ion-induced reactions.

4. Study of prompt radiation emitted in $\text{LCP}$ accompanied fission

Emission of K x-rays in $\text{LCP}$ accompanied fission was investigated for the first time at Trombay where the yields of K x-rays emitted by fission fragments in $\text{LCP}$ accompanied fission were compared with that in binary fission in the case of spontaneous fission of $^{252}$Cf. The observed K x-ray yields in these two cases did not seem to favour the assumption of exclusive emission of $\text{LRA}$ from either of the two fragment groups (Kapoor et al 1968). If the emission of $\text{LRA}$ takes place after scission at the expense of nucleons only from some fragments the isotopic yield distribution of the fragment of a given charge would vary in a definite manner as compared to that in binary fission. With this in view the isotopic yield distribution for fragments of specified nuclear charge in $\text{LRA}$ fission relative to that in binary fission was measured in spontaneous fission of $^{252}$Cf using the method of discrete $\gamma$-ray lines emitted from fragments. A number of $\gamma$-ray lines were assigned to specific fission fragments and the ratios of intensities of these $\gamma$-ray lines in $\text{LCP}$ fission and binary fission were determined. Taking $2^+ \rightarrow 0^+$ transition intensities of the even-even nuclei to be a measure of their yields, the yields of several individual fission fragments of known mass and charge in $\text{LCP}$ accompanied fission relative to that in binary fission were obtained (Ajitanand et al 1975). Figure 6
shows the isotopic yield distributions for different fragment charges 
\((Z = 40–60)\) in LCP accompanied fission derived from this experiment and these are 
compared with those in binary fission. The isotopic mass distributions were observed to 
be altered in the light and heavy fragment groups almost to the same extent. If LCP were 
emitted exclusively from either of the fragment groups (light or heavy), substantially 
different shifts in the isotopic yield distribution would be expected. Thus these results 
also supported the view that LRA are emitted at the expense of nucleons from both the 
light and heavy fragment groups or from the fissioning nucleus as a whole.

To sum up, the investigations of LCP emission in fission briefly reviewed above, have 
provided information on the various facets pertaining to the mechanism of emission of 
these particles and have also proved to be useful in exploring the dynamics of the last 
stages of the fission process.

References

Fraenkel 1967 Phys. Rev. 156 1283
Development of thermal reactor lattice physics methods

B P RASTOGI and H C HURIA
Theoretical Physics Division, Bhabha Atomic Research Centre, Trombay,
Bombay 400085

Abstract. This paper describes the reactor physics methods developed for thermal reactor lattices in India. The formulation of the models introduced is based on energy-dependent integral neutron transport theory. More emphasis was put on the development of lattice cell calculational methods which indeed form the basis of physics design of nuclear reactors. The physical formulation and the cross-sections used were subjected to comprehensive validation tests through analyses of experimental information. The comparison of computed and measured parameters have amply brought out the soundness of the physical formalism and the cross-sections used in our calculational procedures.

Keywords. Neutron transport equation; collision probability methods; lattice cells: multi-group cross-sections; light water lattices; heavy water lattices; rod-clustered lattices.

PACS No. 28:40

1. Introduction

Evaluation of neutron economy is the basic problem in establishing the feasibility of a self-sustaining chain reaction in a multiplying medium. The basic question to be answered in this context is, "how many neutrons are emitted in fissions and what fraction of these can be expected to induce further fissions without being lost in some other reactions". The answer will essentially involve detailed accounting of various mutually competing neutron events depending upon the material composition and geometrical configuration of the system.

The most obvious way of solving this problem is the direct numerical integration of the energy-dependent neutron transport equation for the system under study. But the number of variables involved coupled with the accuracy requirement make this task an impracticable one even on most powerful computers. This naturally leads one to seek simplifications. A number of techniques have thus been developed since about 1940. Better understanding and accurate prediction of reactor behaviour were always stressed upon in the physical models evolved. The three well-established ways of describing neutron events in a thermal reactor are: (i) descriptive set of formulae—the celebrated four factor formula with one or two group leakage model; (ii) few group diffusion or transport theory methods; and (iii) detailed description of all neutron processes by multi-group transport theory.

In this paper, we will be concentrating mainly on the third approach and discuss the methods evolved by us in this area. After some generalities on the integral transport theory, we will return to the methods developed by us in this subject.

In the transport theory approach, the reactor behaviour is predicted through a
detailed assessment of various physical processes involving neutrons. This entails an accurate knowledge of neutron flux distribution as a function of space and energy. Multigroup transport theory methods offer an immediate solution because, both in principle and in practice, these methods are capable of giving a high degree of accuracy and reliability.

The energy variable of neutron flux distribution is discretised in the multigroup approach, that is, the entire energy range is divided into a finite number of discrete energy groups. The selection of the number of groups and their boundaries may be decided by the energy behaviour of the important cross-sections and also by the operating energy range of the system under consideration. For instance, events in the thermal energy range play an important role in thermal reactors and so more weightage may be given to this energy region.

The group cross-sections are obtained by averaging the basic nuclear data with appropriate weighting spectrum. This introduces certain arbitrariness into the group cross-sections because both weighting flux and the group constants are interdependent. Multigroup structure and the associated cross-sections adopted in our model will be discussed in §4. The last section presents the results of a comprehensive evaluation of our model through analyses of a wide base of experimental information.

2. Integral transport equation

There are two alternatives for treating the spatial variable of the transport equation. The integro-differential \((P_n\text{ and } S_n)\) methods and integral transport theory \((P_i; \text{ and } J^±)\) methods (Huria 1976). The former, though mathematically more sound, are rather time-consuming and are thus not suitable for repeated calculations. They are, in fact, more useful for obtaining an overall reactor behaviour.

By contrast, collision probability methods based on integral transport theory have proved to be very effective in solving even the complex power reactor lattice cell problems at relatively low computational costs. They are especially suitable for computing cell flux distribution. We will now devote some time to the integral transport theory and its application to lattice cell calculations.

The time-independent transport equation with isotropic scattering may be written in the form (Williams 1966).

\[
\Sigma_t(r, E)\Phi(r, E) = \int dr' \left\{ S(r', E) + \int_0^\infty \Sigma_g(r', E' \rightarrow E)\phi(r', E')dE' \right\} P(r' \rightarrow r, E),
\]

where \(\phi(r, E)\) is the angle-integrated neutron flux and

\[
P(r' \rightarrow r, E)dr = \frac{\Sigma_t(r, E)}{4\pi |r' - r|^2} \exp \left\{ - \int_0^{|r' - r|} \Sigma_t(s, E)ds \right\}dr
\]

is the probability that a neutron of energy \(E\) originating at the point \(r'\) will make its next collision in the volume element \(dr\) at \(r\). In terms of the transport kernel \(T(r' - r, E)\) defined by
Thermal reactor lattice physics

\[ T (r' \rightarrow r, E) = \frac{\exp \left\{ - \int_{0}^{[r'-r]} \Sigma_s(s, E) ds \right\}}{4\pi |r' - r|^2}, \]  

(3)

\[ P(r' \rightarrow r, E) \text{ may be written as} \]

\[ P(r' \rightarrow r, E)dr = \Sigma_s(r', E)T (r' \rightarrow r, E)dr, \]  

(4)

\[ T \text{ is the probability that a neutron of energy } E \text{ originating at } r' \text{ will reach } r \text{ without making any collision in between.} \]

\[ S (r, E) \text{ is the number of neutrons born per unit volume due to sources. } \Sigma_s(r, E' \rightarrow E)dE \text{ is the cross-section that transfers neutrons of energy } E' \text{ to the energy interval } dE \text{ around } E \text{ at the point } r. \text{ Other symbols have their usual meaning.} \]

The transport equation can be derived directly using the definitions of the transport kernel \( T \) and energy transfer cross-section \( \Sigma_s(r, E' \rightarrow E) \) and considering the neutron conservation.

Let us now specialise the transport equation to the cell problem. The lattice cell under consideration is divided into certain number of (say \( N \)) regions each with constant cross-sections. The equation then becomes (for the flux in the \( i \)th region)

\[ \phi_i(r, E) = \sum_{j=1}^{N} \int_{V_j} dV' T (r' \rightarrow r, E) \left[ S_j(r', E) + \int_{0}^{\infty} dE' \Sigma_{sj}(E' \rightarrow E) \phi_j(r', E') \right] \]  

(5)

where the indices \( i \) and \( j \) refer to individual regions related to \( r \) and \( r' \), respectively.

From here onwards, we will talk only about the collision probabilities and not transport kernel. In terms of collision probabilities, the above equation becomes

\[ \Sigma_{ii} \phi_i(r, E) = \sum_{j=1}^{N} \int_{V_j} dV' P (r' \rightarrow r, E) \left[ S_j(r', E) + \int_{0}^{\infty} dE' \Sigma_{sj}(E' \rightarrow E) \phi_j(r', E') \right]. \]  

(6)

The average flux in region \( i \) is defined by

\[ \phi_i(E) = \frac{1}{V_i} \int_{V_i} dV' \phi_i(r, E). \]  

(7)

The assumption of constancy of cross-sections in each spatial region has enabled us to replace the complete integral by the sum of a series of integrals. If we are interested in the total number of collisions at energy \( E \) in a region \( i \), then it will be necessary to integrate (6) over all points \( r \) in region \( i \) giving

\[ \Sigma_{ii}(E) \phi_i(E) V_i = \sum_{j=1}^{N} P_{ji}(E) V_j \left[ S_j(E) + \int_{0}^{\infty} \Sigma_{sj}(E' \rightarrow E) \phi_j(E')dE' \right] \]  

(8)

where

\[ P_{ji}(E) = \frac{\int_{V_i} dV' \int_{V_j} \phi_j(r', E) P(r' \rightarrow r, E)dr'}{\int_{V_j} \phi_j(r', E)dr'} \]  

(9)
\[ P_{ji} \] is the probability that a neutron born at energy \( E \) through a collision in region \( j \) will make its next collision in region \( i \).

Equation (8) represents a set of \( N \) coupled integral equations for the average flux at energy \( E \) in each region of the system. No approximations have been made so far and the equation in this form is exact and rather general.

The solution of this set of equations evidently requires the specification of collision probabilities \( P_{ij}(E) \). These are dependent on the flux distribution as is clear from their defining equation (9). The most widely used assumption is that of flat flux in each region at all energies. This enables the collision probabilities to be calculated rather easily. With the flat flux (and source) approximation, the set of equations (8) reduces to

\[
\sum_{i} (E) \phi_i(E) V_i = \sum_{j=1}^{N} P_{ji}(E)[S_j(E) + \int_{0}^{\infty} \Sigma_{sj}(E' \rightarrow E) \phi_j(E') dE']
\]

with collision probabilities given by (c.f. (2))

\[
P_{ji}(E) = \frac{\Sigma_{ti}}{4\pi V_j \int_{V_j} \int_{V_i} \exp \left(-\int_{0}^{r'-r} |r'-s| \Sigma_i(s, E) ds \right) / |r'-r|^2}.
\]

The reciprocity theorem which states that the flux at \( r' \) due to a source at \( r \) is equal to the flux at \( r \) due to a source at \( r' \) leads to the following reciprocity relation for the collision probabilities:

\[
\sum_{i} (E) V_i P_{ij}(E) = \sum_{j} (E) V_j P_{ji}(E),
\]

and finally, if the system is effectively infinite in extent, then we must have

\[
\sum_{j=1}^{N} P_{ij}(E) = 1 \text{ for all } i,
\]

because a neutron born in region \( i \) must make a collision either in the same region or in any of the other regions of the system.

Thus, the equation for mean reaction rate in region \( k \) reduces to

\[
\phi_k \Sigma_{ik} V_k = \sum_{i=1}^{N} P_{ik}(E)[S_i(E) + \int_{0}^{\infty} \Sigma_{si}(E' \rightarrow E) \phi_i(E') dE'].
\]

One can rewrite the neutron balance equation for the average reaction rate in region \( k \) in terms of volume and surface sources. In the frame of multiregion picture, the surface source integral is expressed in terms of partial surface integrals, and the volume source in subregion integrals (Emendorfer 1973). Assuming that all the partial currents have the same angular distribution at each surface, each region gets coupled to its adjacent ones through incoming currents. The balance equations could thus be written as:

\[
\phi_k (E) \Sigma_{ik} V_k = V_k Q_k P^k_{\nu\nu} + P^k_{iv} J^+_{k-1} + P^k_{ov} J^+_{k} + P^k_{io} J^-_{k-1} + P^k_{o0} J^-_{k}
\]

\[
J^+_k = V_k Q_k P^k_{\nu\nu} + J^+_{k-1} P^k_{iv} + J^-_{k} P^k_{ov} + J^-_{k-1} P^k_{io} + J^-_{k} P^k_{o0},
\]

where \( P^k_{ix} \) are the first flight collision probabilities with \( x \) and \( y \) indicating the origin and exit processes respectively. For example, \( P^k_{\nu\nu} \) is the probability that a neutron born in
region $k$ will make its next collision in the same region; $P_{i(0)v}^k$ is the probability that a neutron entering region $k$ from its inner (outer) surface will collide within $k$. $J_k^\pm$ are the total outward (+) and inward (−) currents at the outer surface of region, $k$. $Q_k$ is the total source which includes both fission and scattering sources, that is,

$$Q_k = S_k(E) + \int_0^\infty \sum_0^\infty (E' \to E)\phi_i(E')dE'.$$

(17)

The first form of the balance equation [equation (14)] is known as the $P_{ij}$ concept. Herein each region is coupled to all the regions of the system by first flight collision probabilities. The second form [equations (15) and (16)] is known as the $J_\pm$ (interface currents) concept where each region is coupled with its adjacent regions by surface currents. Both the concepts are equivalent as far as the correct space and angular-distributions for source and current densities are involved. But the latter form is simpler to handle from the viewpoint of computer storage and time requirements. The additional assumption that all the partial currents have the same angular distribution, which helped to realise this simplicity, will definitely entail a loss in accuracy. However, the advantage accruing from the savings in computational cost far outweighs the marginal loss in accuracy.

In the multigroup form, the two sets of equations may be written as

$$\sum_{tk}^{\overline{g}} \phi_k^{g} V_k = \sum_{k}^{\overline{g}} V_j Q_j^g P_{tk}^g,$$

(18)

with

$$Q_j^g = S_j^g + \sum_{g'=1}^{g} \sum_{s_j} \phi_j^{g'},$$

(19)

$$\sum_{tk}^{\overline{g}} \phi_k^{g} V_k = V_k Q_k^g P_{tk,\overline{g}V}^g + J_{k-1}^g P_{k,\overline{g}V}^g + J_k^g P_{k,\overline{g}0}^g,$$

(20)

$$J_k^+ = V_k Q_k^g P_{k,\overline{g}0}^g + J_{k-1}^g P_{k,\overline{g}0}^g$$

(21)

where the superscript $g$ indicates the energy group and index $k$ denotes the spatial region $k$.

3. Multigroup transport theory methods developed in India

Let us now review our work in connection with the development of multigroup methods for thermal reactors (Huria 1976). For obvious reasons we selected the $J_\pm$ concept—commonly called the interface current approach—for the lattice cell calculations of thermal reactors, both light water and heavy water moderated. The lattice cell calculations in fact form the basis of physics design of nuclear reactors because they provide the necessary input for global reactor calculations. Moreover, they have independent importance as well—they provide information on the fuel element, reaction rate and power, the isotopic depletion and build-up, and for the interpretation of post-irradiation burn-up measurements.

3.1 Cell calculation

The physical formalism of our lattice physics code murli (Huria 1978, 1981) is briefly described below. The balance equation for average reaction rate in region $k$ bounded by
surfaces \( k - 1 \) and \( k \) may be written as (dropping the group index for the sake of simplicity).

\[
\sum_{tk} \phi_k V_k = Q_k P_k^{V} + J_k^{+} P_k^{-1} + J_k^{-} P_k^{0V}
\]  

(22)

where \( Q_k \) is the total source, that is, the number of neutrons (of energy group \( g \)) born in region \( k \); \( J_k^{\pm} \) is the total outward (superscript +) and inward (superscript -) at the \( k \)th surface; \( P_k^{V} (P_k^{0V}) \) is the probability that a neutron entering region \( k \) from its inner (outer) surface will make its next collision within the region \( k \); and \( P_k^{V} \) is the probability that a neutron born uniformly and isotropically in region \( k \) will collide therein again.

The above equation must be supplemented by the conservation equations for neutron currents at the two surfaces. Equations for the outgoing currents are

\[
J_k^{+} = P_k^{0V} Q_k + P_k^{0V} J_k^{-} + P_k^{0V} J_k^{-1},
\]  

(23)

\[
J_k^{-1} = P_k^{0V} Q_k + P_k^{0V} J_k^{-}.
\]  

(24)

Physical interpretation of these equations is quite obvious. The source term is given by

\[
Q_k^g = \left[ \sum_{g'=1}^{G} \sum_{sk} \phi_k^g V_k + \frac{1}{\lambda} \sum_{g'=1}^{G} V \sum_{f k} \phi_k^g \right] V_k
\]  

(25)

where \( \lambda \) is the eigenvalue.

If we separate the source term into a self-scattering part and the remaining contribution, \( Q_k^g \) may be written as

\[
Q_k^g = \sum_{sk} \phi_k^g V_k + Q_k^g',
\]  

(26)

with

\[
Q_k^g' = V_k \left[ \sum_{g'=1}^{G} \sum_{sk} \phi_k^g + \frac{1}{\lambda} \sum_{g'=1}^{G} V \sum_{f k} \phi_k^g \right].
\]  

(27)

The equation for the collision rate then becomes

\[
\sum_{tk} \phi_k^g V_k \left( \sum_{sk} \phi_k^g V_k + Q_k^g' \right) P_k^{V} + P_k^{0V} J_k^{-} + P_k^{0V} J_k^{-1}
\]

\[
= Q_k^g P_k^{V} + P_k^{0V} J_k^{-} + P_k^{0V} J_k^{-1}
\]  

(28)

Let us now define

\[
T_k = \sum_{sk} \sum_{tk} (\text{number of secondaries per collision}),
\]  

(29)

\[
W_k^{V} = P_k^{V} (1 - T_k P_k^{V}),
\]  

(30)

\[
W_k^{0V} = P_k^{0V} (1 - T_k P_k^{0V}),
\]  

(31)

\[
W_k^{0V} = P_k^{0V} (1 - T_k P_k^{0V}).
\]  

(32)
With the introduction of $W_k^{xy}$'s (28) reduces to

$$\sum_{ik} \phi_i V_k = W_k^{xy} Q_k + W_k^{xy} J_k^- + W_k^{xy} J_{k-1}^+.$$  \hspace{1cm} (33)

One can similarly derive the conservation equations for currents. These are, finally

$$J_k^+ = W_k^{xy} Q_k + W_k^{xy} J_k^- + W_k^{xy} J_{k-1}^+,$$  \hspace{1cm} (34)

$$J_{k-1}^- = W_k^{xy} Q_k + W_k^{xy} J_k^- + W_k^{xy} J_{k-1}^+,$$  \hspace{1cm} (35)

with

$$W_k^{xy} = P_k^{xy} (1 - T_k P_k^{xy}),$$  \hspace{1cm} (36)

$$W_k^{yi} = P_k^{yi} (1 - T_k P_k^{yi}),$$  \hspace{1cm} (37)

$$W_k^{yp} = P_k^{yp} + P_k^{yp} T_k P_k^{yp} (1 - T_k P_k^{yp}),$$  \hspace{1cm} (38)

$$W_k^{yi} = P_k^{yi} + P_k^{yi} T_k P_k^{yi} (1 - T_k P_k^{yi}),$$  \hspace{1cm} (39)

$$W_k^{yi} = P_k^{yi} T_k P_k^{yi} (1 - T_k P_k^{yi}).$$  \hspace{1cm} (40)

The $W_k^{xy}$ are related to the first flight collision probabilities as shown above. They can also be interpreted as multiple collision and escape probabilities as discussed by Driggers (1964). The superscripts $x$ and $y$ denote the origin and exit processes which may be (i) collision event [$V$], (ii) entering or leaving through inner surface [$I$], and (iii) entering or leaving through outer surface [$O$]. For instance $W_k^{xy}$ represents the number of collisions caused in regions $k$ by one neutron entering from its outer surface. Similarly, $W_k^{yi}$ is the number of neutrons leaving region $k$ from its inner surface per neutron that enters it through its outer surface. The neutrons leaving result either from no, one or more collisions that the incoming neutron might undergo in the region.

It is seen that the cosine current approximation leads to balance equations of block diagonal form, each block representing a 2 x 2 matrix of incoming and outgoing currents at each surface. This feature results in substantial reductions in computer storage requirements.

These equations could be solved by appropriate matrix manipulations. However, a slightly different technique is adopted in the present model (Driggers 1964).

The currents at the $k$th surface are expressed in terms of those at the $(k-1)$ as follows:

$$J_k^+ = A_k J_{k-1}^- + B_k,$$ \hspace{1cm} (42)

$$J_k^- = D_{k+1} J_k^+ + H_{k+1}.$$ \hspace{1cm} (43)

Substitution of these into the current conservation equations [(34) and (35)] leads to recurrence relations for the constants $A_k$, $B_k$, $D_k$ and $H_k$. With a specified boundary condition (say at the outermost region $N$ of the cell) like $J_N^- = \beta J_N^+$ or $J_N^- = \text{constant}$, all the relational constants could be evaluated.

Now, the only conservation equation for the centremost region is

$$J_1^+ = W_{1}^{yi} Q_1 + W_{1}^{yi} J_1^-,$$ \hspace{1cm} (44)

which could be reduced to

$$J_1^+ = (W_{1}^{yi} H_2 + W_{1}^{yi} Q_1) / (1 - D_2 W_{1}^{yi}),$$ \hspace{1cm} (45)

$$J_1^- = D_2 J_1^+ + H_2.$$ \hspace{1cm} (46)
With the starting currents given by (45) and (46), the currents at the other surfaces can be easily derived using relations [(42) and (43)].

The first flight collision probabilities as desired for the evaluation of $W_k^{xy}$'s are calculated from the expressions proposed by Kennedy (1964).

3.2 Leakage calculation

The cell calculations give directly the infinite multiplication factor $K_\infty$. The effective multiplication factor $K_{\text{eff}}$ is obtained by representing the leakage normally through experimental buckling. This facilitates the leakage calculation to be done by point model, that is, zero-dimensional multigroup diffusion theory with cell homogenized cross-sections. This procedure was verified by performing multigroup one-dimensional diffusion theory calculations for some representative cases. This latter procedure, however, becomes desirable for relatively small systems having large neutron leakages.

Besides the limitations of the point model, two other effects that have bearing on leakage treatment are the scattering anisotropy and heterogeneity correction. The present model does not account for the former explicitly except for the use of transport corrected total cross-sections. The latter effect for the regular light water lattices is not likely to be significant, because of the fact that the transport cross-sections of light water and UO$_2$ are not too different.

3.3. Burn-up calculations

Burn-up prediction is another important aspect of physics design of reactors. It supplies information on isotopic changes and their effect on reactivity, power distribution, kinetic characteristics, etc. The main outcome of the lattice cell burn-up calculations is the energy and space-averaged constants at desired exposures as required for the global reactor calculations. This will obviously involve the recomputation of lattice cell spectrum at the specified burn-up values.

The isotopic build-up and depletion equations of fissile, fertile and 'futile' (fission product) nuclei are written as

$$\frac{dN_i}{dt} = [\Phi \Sigma_f \gamma_{ij} \bar{\sigma}_f N_j + \Phi \bar{\sigma}_e N_{i-1} + \lambda_i N_i']$$

- $[\Phi \bar{\sigma}_{al} N_i + \lambda_i N_i ']

where $N_i$ = number density of the $i$th type of nuclei; $\gamma_{ij}$ = fission yield of the $i$th nuclei from the fission of $j$th nuclei; $\bar{\sigma}_e$ = average capture cross-section; $\bar{\sigma}_f$ = average fission cross-section; $\bar{\sigma}_a = \bar{\sigma}_e + \bar{\sigma}_f$ = average absorption cross-section; $\lambda$ = decay constant and $\Phi$ = average flux.

The terms within the first pair of square brackets are rates of production of $i$th nuclei from (i) fission, (ii) neutron capture and (iii) decay from other nuclei, respectively. The terms within the second pair are the losses by (i) neutron absorption, and (ii) decay to other nuclei.

These equations are in reality non-linear, because the flux, $N_i$'s and $\bar{\sigma}_i$'s are time-dependent. Assuming the neutron flux to be constant over the given time interval, the coefficients of $N_i$ become constant and the system of equations becomes linear within that time interval. This makes the integration completely straightforward. Thus the generation of burn-up dependent parameters will imply periodic recomputation of the
coefficients of the aforesaid equations using the up-dated neutron flux corresponding to a new isotopic composition.

The total number of burnable (plus fission product nuclei) in our burn-up model is 18 which includes 4 pseudofission products emanating from the chains of uranium and plutonium, rhodium, xenon and a lumped fission product. The remaining are fissile and fertile nuclei. These equations are optionally solved by either 5th order Runge-Kutta method or by trapezoidal rule.

4. Multigroup structure and cross-sections

Energy variation of the neutron spectrum is calculated by the standard multigroup procedure. The group structure and the cross-sections of the \textsc{wims} (Askew 1966) code have been adapted in our calculational model. The main aim in selecting the group boundaries of \textsc{wims} has been to provide enough groups so that the group constants may be considered effectively independent of the weighting spectra used to generate them.

The entire energy range is divided into 69 groups. The fast range defined to lie between 10 MeV and 9-118 keV is divided into 14 groups of equal lethargy width. The limits of 13 resonance groups (9-118 keV-4 eV) have been selected to give adequate representation to the locations of important resonances of main fissile and fertile isotopes. Forty two groups describe the thermal range below 4 eV. Here the group boundaries have been chosen to cluster around the thermal energy resonances of plutonium isotopes and around the peak of thermal spectrum.

The basic cross-sections have been derived from \textsc{uk Nuclear Data Files} and then processed into multigroup constant through \textsc{galaxy} (Bell 1964) code. The weighting spectrum for producing fast cross-sections is that typical of an undermoderated light water system. The cross-sections have been checked to reproduce measured age in the principal moderators.

The cross-sections in the resonance range are not only spectrum-dependent but also change with temperature and composition of the system. The library contains tabulations of group resonance integrals for $^{234}$U, $^{235}$U, $^{236}$U, $^{238}$U and $^{239}$Pu in terms of potential scattering and temperature. These have been derived from the numerical integration of the slowing down equation in the homogeneous media of the absorber and a moderator at different temperatures through the \textsc{sdr} (Brissenden and Durston 1965) code. For actual heterogeneous cell calculations, the effective potential scattering cross-section is obtained through “equivalence theorems” according to the prescriptions of \textsc{wims}.

Thermal scattering matrices derived from various scattering kernels for the usual moderating materials have been tabulated at typical reactor temperatures. All other materials have gas law scattering matrices. The weighting spectra used to obtain other cross-sections are Maxwellian with $1/E$ tails. These have been generated by the code \textsc{pixse}.

The \textsc{wims} cross-sections (and procedures) have been subjected to systematic checks through comparisons with measured parameters and results from sophisticated Monte-Carlo calculations (Fayers 1967; Kemshell 1972; Chawla 1972).

Apart from the 69-group library, condensed 28 and 27 group libraries have been generated using typical operating BWR and HWR spectra for the routine calculations for light water and heavy water power reactors. The condensed libraries use the 69-group library as the basic cross-sections.
An exercise to generate resonance group cross-sections for \(^{235}\text{U}\), \(^{238}\text{U}\) and \(^{239}\text{Pu}\) has also been carried out using a simplified procedure of integrating the slowing down equation by multigroup technique. The fine group cross-sections were prepared using the data from ENDF/B-IV. The group resonance integrals calculated in this way agreed very well with those of WIMS library. The resonance cross-sections for \(^{232}\text{Th}\) and \(^{233}\text{U}\) were generated by this procedure using the basic data from ENDF/B-IV. These have been incorporated into the library. Moreover, the cross-sections of main fissile and fertile materials—\(^{235}\text{U}\), \(^{238}\text{U}\), \(^{239}\text{Pu}\), \(^{240}\text{Pu}\), \(^{241}\text{Pu}\) have also been generated and compared with the original library.

5. Evaluation

The general objective in the development of MURLI code was to acquire an efficient and reliable calculational tool for power reactor design studies. The adequacy and reliability of a calculational model could be established by analysing a wide variety of relevant experimental results. In this context, the measurements in clean uniform lattices provide a sound base for evaluating a code. The model should be able to predict accurately the properties of simple uniform systems. If it cannot do so, it is unlikely to give reliable results for the operating power reactors, which, in reality, have complex lattice structure.

Information for a large number of critical and exponential experiments was collected for investigating the validity of the physical model of MURLI and its associated cross-sections. The list includes not only \(^{235}\text{U}\) enriched systems, but mixed oxide (PuO\(_2\)-UO\(_2\)) and (\(^{232}\text{Th}-^{233}\text{U}\)) oxide experiments as well. Besides the recommended benchmarks (CSEWG 1978), the selected experiments cover a broad range of parameters of importance in physics design aspects like moderator-to-fuel volume ratio, fuel pellet size, fuel enrichment (and isotopic composition), and buckling, etc.

We will begin with the analyses of light water moderated systems. The investigations include not only the calculation of reactivity but also the comparison of measured and calculated reaction rate ratios (wherever available). The following two sections present the results of extensive validation tests carried out for the selected experimental data on critical and exponential experiments moderated by light water.

Part I: Light water lattices

5.1 The effective multiplication factor. The calculation of reactivity is the most fundamental problem of all reactor design studies. The calculational model should be able to predict accurately this important characteristic of the reactor system. All the other parameters like reactivity coefficients, control rod worth fuel depletion effects, etc. are translated in terms of changes in the multiplication factor. MURLI was used to calculate the effective multiplication factor for a large number of clean uniform lattice experiments, the results of which are discussed below (Huria 1981).

(i) U-metal lattices: 44 BNL (Price 1966) exponential experiments using enriched uranium metal rods were analysed by MURLI. The enrichment varied from 1 to 1-3 w\% and moderator-to-fuel-volume ratio from 1 to 4. The mean and the standard deviations of predicted \(K_{\text{eff}}\) were 0.9981 and ±0.00477.

(ii) UO\(_2\) experiments: 55 enriched UO\(_2\)-rod fuelled, light water moderated critical and
exponential experiments (Bernocchi 1977) with a comparatively wider range of enrichment (1-3 to 5-7 w%) and H$_2$O to UO$_2$ volume ratio (1 to 5) were analysed. The mean and the standard deviations of the computed eigenvalues for all the cases considered together were 0.9974 and ±0.00578.

(iii) **Mixed oxide experiments**: The plutonium systems are theoretically more difficult to calculate than the uranium ones. This is due to the complex cross-section behaviour of plutonium isotopes. In particular, $^{239}$Pu and $^{240}$Pu have large thermal energy resonances leading to relatively more significant thermal heterogeneities. Moreover, the fission cross-section of $^{239}$Pu has a non $1/v$ behaviour. Another inherent (and rather undesirable) characteristic of nearly all the plutonium lattice experiments is their small size and consequently a larger fraction of neutrons leaking out. Thus, the predictive capability of a calculational model for mixed oxide system will present a very severe test for its formulation and the cross-sections.

With the above object in view we selected a set of 53 experiments covering wide spectrum of relevant physics design parameters. The discrepancies in the calculated $k_{eff}$'s were within ±6 mk (0.6% of unity).

(iv) ($^{233}$U$-^{232}$Th$_2$) lattices: There are not many experiments using ($^{232}$Th$-^{233}$U) bearing fuel. However, the BNL (Windsor 1970) exponential experiments with (3% $^{233}$U$-97$% $^{232}$Th$_2$) fuel have been frequently used as experimental benchmarks for evaluating both the computational techniques and the data. We have also used the same experiments for testing the adequacy of MURLI for predicting the parameters of ($^{233}$U$-^{232}$Th) oxide bearing fuels. The MURLI calculations for $k_{eff}$ are quite satisfactory and are in reasonable agreement with results from WIMS and Monte-Carlo.

5.2 **Comparison of reaction rate ratios.** The global situation represented by the effective multiplication factor could be satisfactorily and reasonably predicted for a wide variety of uniform critical and exponential experiments. The performance of the theoretical treatments for various neutronic events separately could be better assessed from the comparison of calculated and measured reaction rates. The recommended benchmark experiments could be best utilised for this purpose.

We have also compared relative reaction rates with the corresponding measured values for a number of experiments including the lattices recommended by the Cross Section Evaluation Working Group. The results were satisfactorily consistent and within experimental uncertainties (Huria 1981).

**Part II: Heavy water lattices**

5.3 **The effective multiplication factor.** For the evaluation of MURLI for predicting the physics characteristics of heavy water moderated lattices, we have selected the compilation of relevant data given by Honeck and Crandal (1964). This set includes only natural uranium metal single rod lattices. There is a very wide range of rod radius, lattice pitch, buckling and D$_2$O purity covered by these lattice experiments. The present analysis includes only reactivity predictions for these lattices. As regards the relative reaction rates, the benchmarks recommended by Cross-Section Evaluation Working Group were considered, the results of which will be discussed in §5.4.

Besides the single rod lattices, rod clustered systems were also considered. The calculational model had to be modified to include the appropriate heterogeneity effects of the cluster. We will come to this aspect in §5.5.
Natural U-metal rod lattices: The mean and standard deviations of the computed eigenvalue ($K_{efr}$) for the 90 single rod lattices are 0.99927 and 0.003666 respectively. This indicates that the formulation and the associated cross-sections are quite adequate in predicting the reactivity of heavy water lattices over a broad spectrum of physics design parameters.

5.4 Comparison of reaction rates. The Cross-Section Evaluation Working Group of BNL has recommended the slightly enriched U-metal fuelled experiments done at MIT as benchmarks for evaluating the nuclear data for the heavy water moderated systems (csewg 1978). We have studied these lattices for evaluating our model (and cross-sections as well).

The effective multiplication factors were seen to be very close to unity, the maximum deviation being 0.57%. But the discrepancies noticed in the comparison of calculated and measured reaction rate ratios were outside the acceptable range of accuracy. Even more accurate simulation of the experimental set-up by two-dimensional finite difference diffusion theory and by finite element method did not lead to any marked improvement in these comparisons. It could be added here that similar observations have been made by other authors also for these benchmarks.

5.5 PHWR cluster experiments. A number of experiments have been carried out in Canada with 7-, 19-, and 28-rod clusters with different coolants. All the clusters have natural UO$_2$ fuel pellets with zircaloy-2 cladding and the lattices are moderated by heavy water (Green 1962, 1964a, b; Serdula 1966; De Lange 1966). These were analysed by two different code systems described below:

Climax—It is an extension of MURLI to treat the PHWR cluster lattices. A ringwise homogenisation of the cluster is adopted for spatial flux distribution calculation. However, the heterogeneity effects in fast and resonance energy ranges are obtained through collision probability treatment of explicit cluster geometry.

Club—The physical model of CLUB (Krishnani 1982) is based on a combination of $P_{ij}$ and $J^\pm$ methods. The interaction between various zones within a ring and their contribution to outgoing currents at ring surfaces are calculated by $P_{ij}$ method, while each region is connected to the neighbouring regions by the $J^\pm$ approach. The angular flux at each region interface is expanded in terms of spherical harmonics to correct for deficiencies of the cosine current approximation.

The general observation emerging from computed results is that the calculations of CLIMAX and CLUB agree consistently with the corresponding experimental values and also amongst themselves.

6. Conclusions

The multigroup integral transport theory provides extremely fast and accurate methods for the analysis of thermal reactor lattices. We have developed a number of computer codes based on these methods. These are being used for the design, analysis and operational problems of our light water and heavy water moderated power reactors.
References

Askew J R et al 1966 J. Br. Nucl. Energy Soc. 5 564
Bernocchi E Martinelli R 1977 Light water lattice data; Report NEACRP-U-190
Brissenden R J and Durston C 1965 The calculation of neutron spectra in the Doppler region; Report ANL-7050
Chawda R 1972 An assessment of methods and data for predicting integral properties of thermal reactor physics experiments; Report AEEW-R-797
De Lange P W 1966 Experimental initial conversion and fast fission ratios for clusters of natural U and UO2 in D2O; Report AECL-2636
Driggers F E 1964 A method of calculation neutron absorptions and flux spectra at epithermal energies; Report AECL-1966
Emendorfer D 1973 Physics assumptions and applications of collision probability methods. Invited talk at the Tropical Meeting on mathematical models and computational techniques for analysis of nuclear systems, University of Michigan April 1973
Fayers F J 1967 J. Br. Nucl. Soc. 6 161
Green R E et al 1962 Highlights of Chalk River work on the physics of heavy water lattices since the 1959 Panel Meeting; Report AECL-1684
Green R E et al 1964a Lattice parameters measurements in ZED-2; IAEA Symposium on exponential critical experiments, Vol. III 457
Green R E et al 1964b Lattice studies at Chalk River and their interpretation; Report AECL-2025
Honeck H C and Crandall J L 1964 The physics of heavy water lattices reactor technology, Selected reviews
Huria H C 1976 Multigroup methods in thermal reactor calculations IT-1 Proc. Symp. on Reactor Physics, Bombay, India
Huria H C 1978 Atomkernenergie 31 87
Kemshell P B 1972 Some integral properties of nuclear data deduced from WIMS analysis of well thermalised uranium lattices Report AEEW-R-786
Kennedy J M 1964 Quoted by Driggers 1964
Price G A 1966 Uranium water lattice compilation, Part-I, BNL exponential experiments, Report BNL-50035
Serdulla K J 1966 Lattice measurements with 28-element UO2 fuel assemblies, Part-I, buckling for a range of spacing with three coolants, Report AECL-2686
Williams M M R 1966 The slowing down and thermalization of neutrons (Amsterdam: North Holland)
Windsor H H 1970 Nucl. Sci. Engg. 42 150
Special features of the safety and control systems of the Dhruva reactor

A K RAY and S N SESHADRI
Reactor Control Division, Bhabha Atomic Research Centre, Bombay 400 085, India

Abstract. The prime requirement of reactor safety combined with the need for high availability of nuclear plants have, in recent years, led to considerable research and development efforts at the Bhabha Atomic Research Centre in the field of reactor safety and control engineering. The areas of special interest have been the development of a fast acting emergency shutdown system, on-line fault detection facility for the reactor protection circuits, enhanced instrumentation capability for measurement of critical plant parameters and computerised systems for plant protection, control, performance evaluation, disturbance analysis, and data acquisition and display with particular attention to the problem of man-machine interface. Some of these recent concepts have been incorporated in safety and control systems of the Dhruva reactor which is at present undergoing commissioning trials at Trombay. The special features of these systems are highlighted in the paper. The safety strategy adopted for the reactor and the consequent development of special safety systems are described in detail. The choice of the reactor control scheme and the methodology followed in the design of the automatic power control system are indicated. Campbell instrumentation for measurement of neutron flux or in other words reactor power, extensive use of microprocessors in safety related instrumentation and an improved man-machine interface through suitable design of control room have helped in achieving a high degree of reactor safety. The salient features of these systems are also included.

Keywords. Reactor safety; reactor control; control and instrumentation.

PACS No. 86.10

1. Introduction

Dhruva is the latest addition to the family of nuclear research reactors in Trombay. Designed, constructed and presently being commissioned by engineers and scientists of the Bhabha Atomic Research Centre (BARC), the reactor will provide a thermal neutron flux of 180 trillion neutrons cm\(^2\)/sec at the rated power of 100 Mw(th). This high neutron flux makes the reactor a versatile nuclear facility for fundamental as well as engineering research, for conducting prototype tests on power reactor fuels and materials, and also for production of a variety of radioisotopes to be used in industry, medicine and agriculture.

Certain special features have been incorporated in the design of safety, control and instrumentation systems of Dhruva. These features, hitherto untried in other reactors of India, help in achieving a high degree of reactor safety and plant availability. The evolution of these new systems and their design features are presented.

This article has been specially prepared for commemorating the sixtieth birthday of Dr. Raja Ramanna, who has been a great source of inspiration and guidance to the Dhruva Project and through this we pay our tributes to this great nuclear scientist.
2. Brief description of the reactor

*Dhruva* is a natural uranium fuelled, heavy water moderated and heavy water cooled reactor. Fuel elements in the form of multipin clusters of uranium metal clad in aluminium are located in specially designed coolant channels made of zircaloy. The stainless steel reactor vessel houses the coolant channels and is placed in a concrete vault filled with demineralised light water. The vault acts as a shield against nuclear radiation emanating from the reactor core. The hot heavy water carrying the heat generated in the fuel rods is cooled by demineralised light water in a set of three heat exchangers. The light water in turn transfers its energy through another set of five heat exchangers to sea water, the sea being the ultimate heat sink. In order to maintain purity of heavy water the free space in the reactor vessel is filled with helium. Reactor power is controlled through suitable adjustment of moderator level. Provision has been made for on-power fuelling of the core through a specially developed fuelling machine. A simplified process flow diagram is shown in figure 1 and table 1 indicates the salient features of the reactor.

3. Safety strategy

The safety strategy adopted for *Dhruva* has evolved out of the general principles of reactor safety normally followed in the design of nuclear power plants.

![Simplified process flow diagram of Dhruva.](image-url)
Table 1. Salient features of Dhruva

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reactor type</td>
<td>Thermal research reactor</td>
</tr>
<tr>
<td>Reactor power</td>
<td>100 MW (thermal)</td>
</tr>
<tr>
<td>Thermal neutron flux</td>
<td>$1.8 \times 10^{14}$ neutrons/cm$^2$·sec</td>
</tr>
<tr>
<td>Average specific power</td>
<td>15.2 kW/kg of uranium</td>
</tr>
<tr>
<td>Fuel assembly</td>
<td>7 pin clusters of natural uranium metal clad in aluminium</td>
</tr>
<tr>
<td>No. of fuel assemblies</td>
<td>129</td>
</tr>
<tr>
<td>Moderator and reflector</td>
<td>Heavy water</td>
</tr>
<tr>
<td>Primary coolant</td>
<td>Heavy water</td>
</tr>
<tr>
<td>Secondary coolant</td>
<td>Light water</td>
</tr>
<tr>
<td>Ultimate heat sink</td>
<td>Sea water</td>
</tr>
<tr>
<td>Shielding</td>
<td>Light water followed by high density concrete</td>
</tr>
<tr>
<td>Reactor control method</td>
<td>Adjustment of moderator level in reactor vessel</td>
</tr>
<tr>
<td>Primary shutdown system</td>
<td>Gravity drop of cadmium shutoff rods in the core</td>
</tr>
<tr>
<td>Emergency shutdown system</td>
<td>Injection of gadolinium solution in in-core tubes</td>
</tr>
<tr>
<td>Additional shutdown system</td>
<td>Dumping of moderator</td>
</tr>
<tr>
<td>Refuelling</td>
<td>On-power</td>
</tr>
<tr>
<td>Reactor block dimension</td>
<td>11.05 metres in diameter</td>
</tr>
<tr>
<td></td>
<td>11.94 metres in height</td>
</tr>
<tr>
<td>Reactor vessel dimension</td>
<td>3.76 metres in diameter</td>
</tr>
<tr>
<td></td>
<td>4.19 metres in height</td>
</tr>
</tbody>
</table>

3.1 General principle

The primary goal of reactor safety is to prevent any significant release of radioactive material from a nuclear facility to the general public outside the plant (Thompson and Beckerley 1964). In order to achieve this goal a defence-in-depth approach is normally adopted for the design of a nuclear reactor. This design approach involves two parallel principles. It is first assumed that a serious accident can take place and, therefore, means must be found to prevent the release of fission products out of the plant. This leads to the construction of multiple barriers against radioactive release like fuel sheath, primary heat transport system, containment etc. The second principle demands that accidents must not take place. This results in specially developed safety systems totally dedicated to accident prevention. These ensure that the reactor is automatically shutdown whenever minor plant upsets tend to become major. These also provide long-term cooling to the fuel elements and contain potential release of radioactivity.

The above features namely, provision of safety systems which prevent accidents in the first place, and multiple barriers to radioactive release which ensure public safety even in the event of a serious accident, have been incorporated in all reactors of India and they are largely responsible for the excellent record of reactor safety maintained in this country. However, constant effort for continuous improvement in the state of the art is an interesting feature of nuclear technology. The search for further improvement in reactor safety leads to the conclusion that defence-in-depth approach should be
further extended to the design of the overall safety system itself. This has been tried in the Dhruva reactor.

3.2 Defence-in-depth in safety system design

Defence-in-depth design approach implies a need to ensure that accidents will be prevented even if any individual safety system fails to perform its safety functions. In order to meet this requirement it will be useful to have a reactor concept which by virtue of its intrinsic characteristics prevents accidents or minimises their consequence. It is also necessary to have sufficient redundancy in the safety systems so that failure of any individual system does not diminish the protection capability of the overall system. Since recent experience shows that identical systems can fail simultaneously due to common-mode faults (Moore and Hanauer 1972), redundant safety systems should be based on diverse principles of operations. Defence-in-depth philosophy also demands that individual safety systems must independently perform their safety functions whenever needed. This requirement of high availability can be largely met by employing fail-safe design philosophy. However, it is to be recognised that an absolute fail-safe feature is nearly impossible to achieve in practice. Preparedness for unsafe failures is, therefore, an important aspect of defence-in-depth safety strategy. An on-line fault-checking facility helps in immediate detection of unsafe failures and this feature, if incorporated, will improve the protection capability of safety systems.

The safety strategy adopted for Dhruva thus departs from the traditional approach in its special emphasis on: (i) intrinsic safety features of the plant, (ii) diversely functioning safety systems and (iii) continuous on-line fault-detection facility.

4. Special safety features

The primary cause of practically all reactor accidents is a mismatch between the energy liberation rate within the core and the energy removal rate of the coolant. Accidents can be prevented to a great extent if reactor power is always restricted within the heat transport capability of the cooling systems. The disturbances which tend to upset the power-coolant balance are many and can be classified under two broad categories. These are: (i) externally or internally induced reactivity transients (ii) failures of the heat transport systems. Since the possibility of a major power-coolant mismatch cannot be ruled out, it is necessary to ensure a positive reactor shutdown during such an occurrence.

For Dhruva the inherent characteristics of the plant as well as the special features incorporated in the reactor shutdown systems help in achieving this goal.

4.1 Intrinsic characteristics

A sudden loss of coolant from the reactor core is the most severe of all credible causes of a major power-coolant imbalance. Although a failure of this kind is highly improbable in a nuclear plant, public safety needs to be ensured even for such eventualities. Loss of coolant accident (LOCA) poses a serious safety problem due to its adverse impact on both reactor power and the heat transport system. This situation not only leads to a total loss of heat transport capability but also tends to increase reactor power due to the resultant positive reactivity effect. It has been estimated in the case of Dhruva that the
Safety and control systems of the Dhruva reactor

Gain in reactivity on a total coolant loss can be as high as 8.7 milliK. Undesirable though it is from the safety standpoint, positive reactivity feedback from a LOCA is an inherent nuclear characteristic of all pressure tube type of reactors. This nuclear limitation has been overcome in the Dhruva reactor by an ingenious engineering design of the primary coolant system.

The primary coolant circuit of Dhruva is divided into three independent loops, each consisting of an operating pump, a heat exchanger, valves and interconnecting piping. Figure 2 shows a simplified flow diagram of a single loop. It may be seen that cool heavy water from the inlet plenum flows upwards through the coolant channels containing the reactor fuel and then joins the outlet header. Each of the coolant loops draws hot heavy water from the header and after cooling in the respective heat exchangers discharges the same to the inlet plenum, thus completing the cycle. A part of the cool heavy water at heat exchanger outlet is diverted to the top of reactor vessel for cooling the top tube sheets and other internal structural assemblies. This heavy water after cooling the internal components of the reactor vessel falls on the free surface of the moderator. Thus at every instant heavy water from the closed main coolant loop enters into the free moderator region of the vessel. In order to maintain dynamic equilibrium, it is necessary to return an equal quantity of heavy water from the moderator to the coolant system. This has been achieved by connecting the reactor vessel to the suction side of the coolant pump in each loop by suitably sized lines. In this arrangement the reactor vessel acts as a surge tank for the coolant and consequently moderator becomes the source of coolant make-up during a LOCA. Thus a break in the coolant circuit automatically increases the moderator return flow and loss of coolant heavy water is continuously replenished by moderator heavy water. It is clear that fuel cooling will continue so long as moderator inventory in the reactor vessel is not completely lost. The integrated coolant-moderator circuit of Dhruva ensures that a loss of coolant is necessarily preceded by a total loss of moderator which results in an automatic reactor shutdown. Also, a rupture in the coolant circuit does not immediately affect the heat transport capability of the cooling system and this provides sufficient time for actuation.

Figure 2. Simplified primary coolant flow diagram (single loop).
of the emergency core cooling system. These two factors enable preservation of fuel integrity even after a double ended rupture of the largest coolant pipe. LOCA in Dhruva, therefore, does not pose any problem of fission product release even within the plant.

4.2 Emergency shutdown system

The provision of two diversely functioning high speed shutdown systems, either of which is capable of causing a rapid reactor shutdown for the entire spectrum of credible accidents, is a special feature of Dhruva. The primary shutdown system utilises the traditional method of inserting a set of neutron absorber elements, called shutoff rods, into the reactor core. There are nine shutoff rods in Dhruva which use cadmium as the neutron absorbing material. The drive mechanisms of these rods are of electromechanical type and follow the popular clutch and drum arrangement. In this design, the rod is attached to a steel wire rope which in turn is wound on a drum. An electromagnetic clutch, if energised, connects the drum to a motor through an irreversible gear train. When the reactor is in operation, the rod is parked above the core region and is held in position by the energised cluth. On a reactor shutdown signal, the clutch is deenergised and the rod drops freely into the core. Although the shutoff rod assemblies are independent of one another, they are identical in design and construction. Thus, in spite of built-in redundancy the primary shutdown system is vulnerable to common-mode failures. To guard against such failures a diversely functioning emergency shutdown system has been provided. This system, specially developed for Dhruva, injects a neutron absorbing liquid solution (poison) into a set of empty tubes located in the reactor core.

4.2a System description The system, shown in figure 3 comprises essentially of:

![Figure 3. Emergency shutdown system.](image-url)
(i) twenty tubes passing through the reactor, termed as poison tubes, and the connecting headers at top and bottom ends; (ii) a 50 litre capacity liquid poison tank containing solution of gadolinium nitrate in demineralised water, (iii) a 60 litre capacity nitrogen tank; (iv) a 20 litre capacity buffer tank; (v) a 400 litre capacity poison storage tank; and (vi) nitrogen cylinders, a pump, connecting pipes and valves.

The poison tank containing liquid neutron absorber is connected to the lower header and also to the nitrogen tank through three fast-acting solenoid valves (scram valves) operating in parallel. The nitrogen tank is maintained at a pressure of 15 kg/cm² (g) by the nitrogen cylinders. The upper header is connected to the buffer tank which in turn is connected to the poison tank through two fast acting solenoid valves (equaliser valves) operating in parallel. Two more fast acting solenoid valves (vent valves) operating in series, connect the poison tank to an air plenum which is always maintained at near atmospheric pressure. These valves play an important role during the poison injection process. Other valves in the system facilitate post-injection operations like dumping, washing and refilling of poison, and they remain in their deenergised fail-safe states prior to and during the injection process.

4.2 Principle of operation When the reactor is operating, the liquid level in the poison tubes is maintained below the reactor vessel by applying suitable nitrogen pressure to the buffer tank and venting the poison tank to the air plenum. In this condition the scram and equaliser valves are in energised close states and the vent valves are in energised open states. On an emergency shutdown signal these valves are deenergised, thus pressurising the liquid in poison tank. This initiates rapid injection of poison solution into the poison tubes. The time required to fill the tubes is approximately 0.8 second. The high velocity liquid front is damped by the compressed nitrogen in the buffer tank which creates a counter pressure on the moving mass. The motion is found to be oscillatory. However, the oscillation is contained well above the active core and as such there is no variation in reactivity.

In order to ensure a high degree of stability of the poison level in the reactor operating mode, a novel feature utilising the principle of pressure feedback has been incorporated in the system. It may be seen from figure 3 that the upper header pressure is fed back to the so-called ‘bubble tube’ through a feedback line. The junction of these lines, marked as point ‘x’ in the figure, is kept at the same elevation as the desired level in the poison tubes. The normal levels in the poison tank and the poison tubes have an elevation difference of 5.4 meters. Thus a differential pressure equivalent to 5.4 meters of solution head is to be maintained between the tubes and the tank. The upper header is, however, kept at a pressure somewhat higher than the above requirement, so as to maintain a steady flow of gas through the pressure feedback line and the bubble tube. A rise in the poison tube level reduces the gas flow rate, thereby increasing the pressure in the poison tubes. This immediately pushes down the liquid to the original elevation. Similarly, a fall in the level increases the gas flow rate. This reduces the pressure in the poison tubes which again restores the original level.

4.2c Post-injection operations During the poison injection process the nitrogen supply system is isolated from the nitrogen tank. After the injection is completed, vent valves are opened. This releases the pressure built up in the system. For a restart of the reactor, the poison from the core is brought back to the storage tank. The poison tubes are then flushed with demineralised water to remove any residual poison that may stick
to the tube walls. After the completion of flushing, the system is once again filled with poison.

4.2.2 Performance under failure  The dynamics of the poison injection process was analysed through a detailed mathematical model which was verified by actual tests on a scaled prototype of the system. The process simulation was carried out on the BESM-6 computer of BARC. Using the model, the behaviour of the system under normal condition as well as under various modes of failure were determined. This study reveals that poison can be injected in less than one second even if two of the three scram valves and three of the four equaliser and vent valves simultaneously fail to operate. Therefore, satisfactory operation of one scram valve and any of the vent and equaliser valves ensure adequate system performance. Since these valves operate on fail-safe principle, the probability of not meeting the above requirement is extremely small and the same is estimated to be less than 1 in $10^6$. Even in the case of three scram valves failing simultaneously, poison is still injected though at a considerably slower rate. The injection time under such failure is approximately 8 seconds. Figure 4 shows the post-injection rise in poison level as a function of time for the normal mode as well as the extreme condition of all scram valves failing.

4.3 On-line fault detection facility

The failures in a safety system are normally classified under ‘safe’ and ‘unsafe’ categories. A safe failure inadvertently actuates a safety system thereby causing a
spurious reactor shutdown. Although such failures reduce plant availability they have no adverse effect on safety. An unsafe failure decapacitates a safety system and, therefore, is of great concern. Periodic functional testing of safety systems helps in reducing the probability of unsafe failures. The fractional dead time \( (D) \) of a safety system which represents the fraction of operating time that the system may remain decapacitated is given by

\[
D = \frac{1}{2} f_u T
\]

where \( f_u \) is the unsafe failure rate and \( T \) is the interval between two successive tests. It is clear that greater the frequency of tests lesser is the probability of an unsafe failure.

Manual testing of systems involves considerable time and effort on the part of the operators and makes it difficult to achieve rapid rates of testing. High speed automatic test facilities are, therefore, required to overcome the problem of unsafe failures. This has been attempted in some of the critical areas of the Dhruva safety system.

The fine impulse test (FIT) facilities of Dhruva continuously search for unsafe failures in the safety logic systems which generate safety signals and actuate safety mechanisms. Each safety function is generated in three independent logic channels and the relevant mechanism is actuated whenever two or more of these channels demand a safety action. The FIT system is based on a 8-bit Intel microprocessor. It may be seen from figure 5 which represents the system block diagram, that fine impulses are injected into each logic channel as simulated input signals. These pulses represent the conditions which will prevail during an emergency. Sufficiently short pulses used for this purpose ensure that the output devices do not respond to these stimuli. The outputs of the logic channels for the simulated inputs are read by the central processing unit (CPU) through digital input data cards and get checked for any possible fault. When faults are encountered, the system analyses the faults to the card level and displays the address of

![Figure 5. Fine impulse test facility.](image-url)
the defective cards on the display panel. The system takes less than one second to check all components under its purview. A watch dog timer and other self-checking facilities have been incorporated in the system to improve its reliability.

The FIT facilities of Dhruva can carry out checks a million time faster than the manual testing rate normally followed in operating plants. Since the probability of a failure is inversely proportional to the frequency of testing, these facilities provide an extremely effective protection against unsafe failure.

5. Control and instrumentation systems

The control and instrumentation (C & I) systems of Dhruva are designed to achieve a high degree of automation in plant operation. This reduces the routine work load of the operating staff, providing increased opportunity to concentrate on unusual occurrences. The C & I systems also assist the operator in high level monitoring of plant performance, thereby enhancing both safety and availability of the reactor.

5.1 Automatic power control system

Reactor power in Dhruva is regulated through suitable variation of moderator level in the reactor vessel. As shown in figure 6, the moderator level is changed by varying the rate of flow of heavy water into and out of the vessel. The speed of the moderator pumps in the inflow line and the opening of control valves in the outflow line are controlled to achieve the required flow variation. This method of reactor control offers certain safety advantages. It enables rapid draining of moderator from the core by opening of the control valves and tripping of the moderator pumps, thereby causing a reactor shutdown. The control system can, therefore, be used as an additional shutdown system of the reactor.

The dual control of both inflow and outflow rates of heavy water is a special feature of Dhruva. This scheme results in low equilibrium flow in the core which restricts the

![Figure 6. Automatic power control system.](image)
flow-induced vibrations of in-core components. This also reduces the power consumption by the pumps. Added to these the large reactivity range of this scheme enhances the control capability of the system.

To achieve high availability and adequate safety, triplicated control channels are provided. Each channel is essentially independent of the other two and controls one valve and one pump. The system is so designed that control capability is preserved even on failure of one complete channel. However, the possibility of such failures is low since the functioning of the various sub-systems is continuously monitored by a number of signal comparator units incorporated for this purpose. The control signal for the pump or the valve is derived through suitable processing of the power demand signal set by the operator, and the feedback signals provided by the neutron power detecting instruments.

The system design was optimised through a parametric study of the system behaviour under various steady state and transient conditions. Detailed mathematical models were developed for this purpose. The reactor was described by the point kinetics equations with six groups of delayed neutrons. Accelerations in fluid flow were calculated through incompressible one-dimensional fluid flow equations. Individual models were also developed for the electronic units, moderator pumps and control valves. The integrated model was found to be highly nonlinear with 29 state variables. The steady state stability of the system which is valid for small perturbations, was evaluated through the linearised version of the model. It was established that the control system was stable with adequate gain and phase margins. The transient performance was, however, analysed with the help of a detailed computer code which included all the non-linearities of the model.

5.2 Wide range monitoring of reactor power

Neutron flux which is proportional to reactor power, is normally measured by processing a neutron detector output in either pulse mode or d.c. mode. In pulse mode individual pulses from the detector are counted and the count rate represents neutron flux or reactor power. This technique is employed for low level flux up to a maximum of $10^6$ n/s. At high flux values individual pulses overlap and the resultant current is measured in the d.c. mode. The d.c. current is proportional to neutron flux and hence, reactor power. In recent years, a third technique for the measurement of neutron flux has been developed. This is based on the noise theories of Campbell. Applying this theory it can be shown that the mean square of the current from a source of random current pulses is proportional to the pulse rate and square of the pulse height, provided that the d.c. portion of the current is cut off by capacitive coupling and the fluctuating a.c. component is used exclusively. Therefore, the mean square of a.c. components of the detector current is proportional to neutron flux. The Campbell technique provides certain advantages. Since only the variations in current are measured, d.c. leakage currents of cables do not result in significant error. Also, the mean square output being proportional to the square of pulse height smaller pulses contribute less to the total signal, thereby improving the gamma discrimination capability of the system. A proper integration of pulse and Campbell techniques enables wide range monitoring of reactor power through a single neutron detector. Such a system has been developed for Dhruva. This system, shown in figure 7, monitors reactor power over a range of 10 decades using a single fission chamber. It has been, thus, possible to use a few detectors and yet have
wide range monitoring capability. A ten decade coverage of reactor power by the regular instrumentation system is a special feature of Dhruva and this enables normal reactor start-up even after a very long shutdown without the help of additional instrumentation.

5.3 Instrumentation for operator support

A number of microprocessor-based systems assist the operator in high level monitoring of plant performance. These enable early detection of minor plant upsets and guide the operator in taking corrective actions. The data acquisition and display system, coolant channel surveillance systems and alarm annunciation system are provided for this purpose. These systems were specially developed to meet the high reliability requirements of reactor instrumentation.

The data acquisition and display system is configured around a 16-bit Intel microprocessor and employs two-colour cathode ray tubes (CRTs) for data display. The system exploits the advantages offered by colour CRT terminals and presents plant data in easy-to-grasp formats like mimics, graphs, tables, bar charts etc. This leads to enhancement of operator's comprehension of plant behaviour and therefore to a better man-machine interaction. The functions carried out by the system include alarm annunciation with detailed instructions for operator action, display of plant status in mimic and tabular forms, system performance calculations and graphic presentation of the trend in various process variables during routine transients like start-up, manual shutdown, raising or lowering of power etc. Facilities are also available for routine or demand logs of plant parameters.

The coolant channel surveillance systems monitor the health of fuel elements in the reactor core. The coolant flow, outlet temperature and radioactivity level in each coolant channel are continuously monitored by three independent microprocessor-
based systems and the relevant information is displayed in the control room. Whenever any of these parameters exceeds the operational limit the system provides an alarm and indicates the address of the affected coolant channel. In the normal mode process status is displayed on a demand by the operator.

The alarm annunciator system provides audio-visual warning to the operator when minor perturbations occur in operating parameters. In order to achieve high reliability a dual computer scheme has been adopted for this system. In this scheme each computer is entirely capable of performing all functions of the annunciator. Only one computer system is normally active and the second one is a standby. In the event of a malfunction of the active computer system the control is automatically transferred by the watchdog to the standby system.

Extensive self-checking features have been incorporated in all computerised systems of Dhruva and they are responsible for the high reliability of these systems. Apart from the reduction in size and cost, these systems offer an additional advantage of flexibility which facilitates system modification during commissioning.

6. Control room

The control room provides all the information and controls required for safe operation of the plant. Considerable efforts were made to evolve an optimum design of the control room, especially the sizing and layout of panels and consoles as well as the arrangement of control and display devices. The main emphasis was on achieving an optimum integration of the operator with the process through appropriate human factor engineering. This was considered to be essential, since many of the abnormal occurrences in the operating plants can be attributed to operator errors resulting from deficiencies in control room design.

In view of the multitude of tasks to be performed from the control room, it is necessary to have functional grouping of different tasks. The Dhruva control room consisting of nine panels and one console having a total length of approximately 30 meters can be subdivided into three functional areas. Routine reactor start-up and shutdown is carried out at the console. Reactor cooling and other process systems which are unaffected by reactor shutdown, are controlled from the central five panels. The extreme panels, two on each side, provide the information which is to be periodically logged by the operator.

In order to facilitate identification of different controls, operational mimic diagrams are extensively used. These diagrams display the flow of a process medium in a system and include the major components like pumps, valves, tanks, etc, which are represented by symbols. The control switches of an equipment are incorporated within the corresponding symbol. These switches are of lighted push button type and thus equipment status is always indicated on the corresponding symbol. This enables the operator to immediately determine the effect of a control action taken by him. As a further aid to the operator the status of relevant process variables are displayed through suitable indicators. Mimic diagrams are structured in different colours, each colour representing a specific system. This helps in highlighting different systems.

Two colour CRTS located on the console display detailed information on the plant status. This facility enables the operator to gather all data from a single instrument. Individual instruments like recorders and meters also provide standby indications.
7. Conclusion

The safety and control systems of Dhruva were entirely designed, developed and fabricated at BARC. The special features of these systems reflect the current trends towards diversified safety systems, increased use of computer power in C & I functions, enhancement of existing instrumentation capability and improved man-machine interaction through advanced control centres. These concepts are likely to be employed in future nuclear plants of our country.

References

Physics and engineering aspects of fast reactor safety

C V SUNDARAM and S M LEE
Reactor Research Centre, Kalpakkam 603 102, India

Abstract. A vast amount of research and development work has been done in recent years to resolve the issues of relevance to the safety of the Liquid Metal cooled Fast Breeder Reactor (LMFBR). Based on the results of this research as well as on the experience gained from the operation of test and prototype LMFBR’s, a certain consensus is emerging on the safety requirements of a modern sodium-cooled large fast power reactor. The paper reviews the fundamental physics and engineering aspects of LMFBR safety with reference to the Fast Breeder Test Reactor (FBTR) now being commissioned at Kalpakkam, and the proposed larger Prototype Fast Breeder Reactor (PFBR). The elements contributing to the inherent safety of fast reactors are recapitulated followed by description of the philosophy of the plant protection system and the use of engineered safeguards to enhance the safety. Finally, the principles used for the containment of radioactivity are discussed.

Keywords. Fast breeder reactors; safety requirements; engineered safeguards; containment of radioactivity.

PACS No. 28 50; 28 45

1. Introduction

The fundamental concerns of nuclear reactor safety are apparent in the basic nuclear fission chain reaction itself. In this reaction, a neutron splits the nucleus of an atom of a fissile material like uranium or plutonium into two smaller radioactive nuclei (fission products). At the same time, a considerable amount of energy in the form of kinetic energy of the fission product nuclei is released, along with nuclear radiations, and several additional neutrons. The released fission neutrons, which are emitted in a spectrum of energies peaking around 2 MeV, can induce fission in further atoms causing a chain reaction, which process is used for heat production in a nuclear reactor. It is immediately seen that the problems of safety specific to nuclear reactors would have their origin in the following:

a) control of the chain reaction and reactor geometry to prevent unchecked growth of neutron population that can cause heat production in excess of the heat removal capability,
b) shielding of the nuclear radiations produced during the chain reaction,
c) containment of the radioactive fission products, shielding of their nuclear radiations, and provision of cooling to remove their radioactive decay energy when the reactor is shut down.

To appreciate the development of concepts in nuclear reactor safety, it is necessary to recall the events that have occurred in the short period of time that has elapsed since the discovery of the fission reaction in 1939. The possibility of using low energy neutrons for fission of U$^{235}$ and the need for high energy neutrons for fission of U$^{238}$ or thorium...
was shown in the same year, as also the existence of fission neutrons and delayed neutrons. This was followed by Fermi attempting and failing to sustain a chain reaction in a light water moderated natural uranium assembly. The brilliant theorising and experimentation of Fermi’s group led to the important decisions to use graphite as moderator as well as to lump the fuel to minimise the loss of neutrons by non-fission reactions in $^{238}$U. This led to the successful atomic pile CP1 which went critical on December 2, 1942. The reactor used 6 tons of pure uranium metal, 50 tons of compressed uranium dioxide powder, 400 tons of graphite and a number of control rods consisting of cadmium sheets nailed to wood. In addition to shim control rods there were cocked safety rods, which could be released to fall into the core automatically, should the neutron intensity as measured by a BF$_3$ gas filled ionization chamber rise too high.

After this historic achievement, several other reactors were constructed in the USA driven by the need for plutonium for weapons. The first enriched uranium, light water moderated reactor was built in 1944. The first atomic bomb (Trinity test) was exploded on 16th July 1945 and was a plutonium device. This was followed 21 days later by the Hiroshima $^{235}$U bomb and a further 3 days later by the Nagasaki plutonium bomb.

Canada achieved its first chain reaction in the heavy water moderated reactor ZEEP in 1945, UK achieved it in the graphite moderated pile GLEEP in 1947, France in the heavy water moderated reactor ZOE in 1948 and India achieved its first chain reaction in the APSARA reactor in 1956.

The first controlled fast chain reaction was in 1946 in the 25 kWt CLEMENTINE fast reactor of USA using plutonium fuel and mercury coolant. The first nuclear electricity production was also from a fast reactor, the enriched uranium fuelled, NaK cooled 1 MWt EBR 1 which went critical in 1951. India achieved its fast chain reaction in the PURNIMA reactor in 1972.

In general, the safety of a reactor is ensured by parallel approaches at different levels providing a ‘defense in depth’. The first level is the design of an inherently safe plant with an adequate control system which can operate with a high degree of reliability. In addition, special emphasis is placed on the quality of materials and workmanship of components, with provision for continuous or periodic inspection of components and sub-systems. The careful design and construction is followed by equally careful operations so as to reduce the probability of accidents to a very low level.

At a second level, safety is provided by means of a comprehensive plant protection system (PPS) which can safely handle a wide range of conceivable abnormal incidents and malfunctions and safely shut down the plant. The PPS includes a variety of instruments and sensors to monitor the state of the plant and to take protective as well as sympathetic control actions to prevent any abnormality leading to an accident.

The third level of safety is provided by engineered safety features such as reactor containment vessel and building, secondary shutdown systems, emergency core cooling system, etc., which limit the consequences of certain highly unlikely accidents, which are postulated to occur in spite of the first and second level safety measures. Finally, a careful choice of site is made, and exclusion areas defined such that routine as well as accidental radioactive releases do not affect the general population.

In the following discussion, we elaborate on these elements of reactor safety with reference to the fast reactor. The discussion will start with physics, control and engineering design features providing the inherent safety of a modern fast reactor, then go on to the main components of the PPS. Finally, the principles of the containment of radioactivity under normal and accident conditions will be discussed.
The 40 MWt/13 MWe loop type Fast Breeder Test Reactor (FBTR) now being commissioned at Kalpakkam and the proposed 500 MWe pool type Prototype Fast Breeder Reactor (PFBR) will be used as examples during the discussion and their safety features will be highlighted vis-a-vis the international approaches to fast reactor safety.

2. The fast reactor

2.1 Thermal and fast chain reactions

The relative probabilities of the production of neutrons by fission in a reactor, and the loss of neutrons by other nuclear reactions as well as escape from the reactor govern the possibility of maintaining a chain reaction. Of all the naturally occurring isotopes of elements, only U\(^{235}\) is fissile and can sustain a chain reaction. U\(^{235}\) has an abundance of only 0.7% in natural uranium, the rest being the isotope U\(^{238}\). However, the U\(^{235}\) content can be artificially increased in enrichment plants to produce enriched uranium. In order to sustain a chain reaction in natural uranium or low enrichment uranium, it is necessary (see for example, Glasstone and Edlund 1952) to reduce the energy of the fission neutrons by means of scattering collisions with suitable light moderator atoms, till the neutrons attain low energies (around 0.25 eV) at which they are in thermal equilibrium with matter. These thermal neutrons are then allowed to cause further fission and continue the chain reaction.

If a chain reaction is to be maintained in a reactor without a moderator, then it is necessary to use high enrichments ranging from 15% (for large reactors) to over 90% (for small reactors). Here the chain reaction is maintained by the fast moving fission neutrons. There are several advantages of the fast chain reaction over the thermal one. The first is that fast reactors are very compact on account of the absence of the large bulk of moderator which is required in thermal reactors. Fast reactors are much more compact than the light water moderated package power reactors used in nuclear powered sea-craft, and have been considered for space craft applications (Hanson 1984). In addition, the fast chain reaction is used in nuclear explosives applications and in pulsed fast reactors providing a large number of fast neutrons for nuclear research (Oakes 1968).

However, the most important reason for the large scale development of fast reactors is the efficient breeding of new fissile material, plutonium and U\(^{233}\), from their respective fertile materials, U\(^{238}\) and thorium. The fertile materials are several hundred times more abundant than U\(^{235}\) and can be converted to fissile material by neutron absorption in a reactor. A breeder reactor is designed such that it generates more fissile material from fertile material than the fissile material it consumes for energy production. Without breeder reactors, it will be possible to use only a small fraction of the uranium resources. Hence all plans for the long-term utilisation of uranium and thorium resources are based on the breeder reactor. While there are designs for thermal breeder reactors using the Th/U\(^{233}\) fuel cycle, for efficient breeding the fast chain reaction is the best, mainly on account of the fact that the number of neutrons released per absorption in fuel is higher for higher neutron energies.

2.2 The LMFBR

There are different kinds of fast reactors and the best developed is the Liquid Metal cooled Fast Breeder Reactor (LMFBR). In this type, heat is generated by the fission chain
reaction in a core consisting of fuel subassemblies supported vertically on a grid plate in a reactor vessel. Each subassembly is a bundle of fuel pins made of stainless steel clad tubes of small diameter (less than 1 cm) containing the enriched fuel pellets. The fertile material in the fuel as well as on top and bottom of the fuel (axial blankets) get converted to fissile material by neutron irradiation. In addition, radially surrounding the core are blanket subassemblies containing additional fertile material for conversion to fissile material.

Water is a neutron moderator, and rapidly decelerates the fast moving fission neutrons to less than the energy suitable for breeding, hence it cannot be used as a coolant. The problem of removing a large amount of heat from the small compact core of a fast reactor by a coolant which does not reduce the neutron energy unacceptably is solved in the LMFBR by the use of liquid sodium. This coolant has excellent heat transfer characteristics and good chemical compatibility with fuel and structural materials. The reactivity of sodium with air or water is taken care of by providing an inert cover gas like argon over free sodium surfaces. In early small experimental LMFBRs, mercury, potassium and a eutectic mixture of sodium and potassium (NaK) have been also used as coolants.

The sodium coolant is pushed by pumps through the core and gets heated. The hot primary sodium is radioactive and is not used to directly produce steam, but rather transfers the heat to secondary sodium through intermediate heat exchangers (IHX). The non-radioactive secondary sodium flows through sodium heated steam generators to produce steam which drives turbo-generators. In the pool type layout, the whole primary coolant circuit along with primary pumps and IHXs is contained within the reactor vessel as shown in figure 1. In the loop type layout, the primary pumps and IHXs are outside the reactor vessel (figure 2) though the whole radioactive primary circuit is contained within the reactor containment building. Both types of layout are prevalent though they have differing safety implications.

The chain reaction is usually controlled by the use of the neutron absorber boron, in the form of boron carbide control rods. Fuel loading and unloading is invariably done under shutdown conditions, from the top of the reactor vessel by means of fuelling machines handling one subassembly at a time.

---

Figure 1. Schematic diagram of a pool type LMFBR.
The main characteristics of FBTR and of PFBR are given in table 1. FBTR is a mixed plutonium-uranium carbide fuelled LMFBR with two primary loops. PFBR is planned to be a pool type LMFBR using either mixed carbide or mixed oxide fuel and incorporating state of the art safety features.

3. Inherent safety features

3.1 Neutron population and multiplication factor

The power produced in a nuclear reactor core is proportional to the rate of fission which is proportional to the product of the fissile atom concentration and the neutron population, \( N \), in the core. The neutron population level arises from a dynamic equilibrium between the generation of neutrons by the fission of fuel atoms and the loss
of neutrons by absorption in fuel and other atoms and by leakage from the core.

A fundamental parameter is the multiplication factor, $k$, which is the ratio of the rate of generation of neutrons to the rate of loss of neutrons. The fission rate and reactor power will hence continuously increase, decrease or be constant accordingly as $k$ is greater than, less than or equal to unity and the reactor is correspondingly said to be super critical, subcritical or critical. In an operating reactor, $k$ can be changed within limits by movement of the control rods which changes the rate of absorption of neutrons. An alternate parameter to $k$ is the reactivity $\rho$ defined as, $\rho = (k - 1)/k$. Study of the normal and abnormal changes to the core reactivity is important for control and safety.

3.2 Prompt neutron lifetime and delayed neutron fraction

The prompt neutron lifetime, $\tau$, is the average time from the 'birth' of a neutron in the fission of a fuel atom to the 'death' of the neutron by absorption or leakage. In this connection, it must be noted that over 95% of the reactions in a fast reactor are scattering reactions which reduce the neutron energy and prolong the neutron life. In a fast reactor, $\tau$ is of the order of $10^{-7}$ sec as compared to the order of $10^{-4}$ to $10^{-3}$ sec for thermal reactors. The average rate of loss of neutrons is $N/\tau$ and the rate of production is $kN/\tau$. Thus the growth of neutron population, for constant $k$, will be exponential with a time constant $T = 1/(k - 1)$ called the reactor period.

Normal control movements make $(k - 1)$ of the order of $10^{-4}$ to $10^{-3}$ so that the period would be of the order of $10^{-4}$ to $10^{-3}$ sec for fast reactors and of the order of a second for thermal reactors. These are quite rapid response times and would have made reactor control problematic, especially in the case of fast reactors. However, as explained below, the presence of delayed neutrons increases the response time and makes it the same for both fast or thermal reactors.

A small fraction $\beta$ (delayed neutron fraction) of the neutrons produced in the fission of a fuel atom are, in fact, emitted subsequent to the fission by decay of certain radioactive fission products (delayed neutron precursors) with mean lives ranging from 0·3 sec to 80 sec. $\beta$ is 0·00204 and the average mean life $\tau$ is 14·6 sec for Pu$^{239}$ and the values differ for different nuclides (Keppin 1965). When delayed neutrons are included, the reactor period can be obtained, when $k$ is near unity, by replacing the prompt neutron lifetime $\tau$ by a weighted mean neutron lifetime $\tau^* = (1 - \beta)\tau + \beta\tau$. $\tau^*$ is of the order of 0·05 sec and hence for $(k - 1)$ of the order of $10^{-4}$ to $10^{-3}$ sec, a period of the order of 50 to 500 sec will result.

There are two more important effects connected with the delayed neutrons:

(i) When a reactor is shut down by insertion of control rods such that $k$ is reduced much below 1, the power at first falls rapidly to a fraction $\beta/(1 - k + \beta)$ of the operating power, then falls gradually towards zero with a period close to the mean life of the longest lived delayed neutron precursor, viz 80 sec. Thus it is not possible to completely shut off the chain reaction immediately.

(ii) When $\rho$ is greater than $0$ by an amount $\beta$, then the neutron population grows at a rate governed by the extremely short prompt neutron lifetime $\tau$ and not by the weighted mean neutron lifetime $\tau^*$. The reactor under such conditions is said to be super prompt critical and very rapid rates of power increase become possible. An important safety consideration is to protect against the possibility of a reactor becoming super prompt critical. On account of the importance of the delayed neutron fraction on the kinetic
behaviour of the chain reaction, $\rho$ is often measured in terms of $\beta$. An amount of reactivity equal to $\beta$ is called a ‘dollar of reactivity’ which makes the reactor super prompt critical.

3.3 Feed-back coefficients of reactivity

One of the important control and safety features of a nuclear reactor core is that the reactivity is dependent on the core temperature. As seen above, when $k$ is changed from 1, by say control rod movement, the neutron population changes and consequently the power $P$ changes. This leads to changes in the core temperature $T$ which, in turn, affect $k$ and thereby the power. Thus, we have a feed-back loop.

When the feed-back change in $k$ opposes the applied change in $k$, then the reactor has a negative feed-back coefficient. In other words, any addition of reactivity increases power which increases temperature, and which in turn decreases the reactivity. Thus, negative feed-back coefficients lead to stability of operation and add to the inherent safety of the core. On the other hand, a positive feed-back coefficient can lead to instability and power runaways. It is to be noted that the core temperature can also be changed by changes in the coolant flow rate or the coolant inlet temperature. Designing the reactor core for all negative feed-back coefficients of reactivity is an important element in the overall safety of the reactor.

Different mechanisms with differing response times of action contribute to the feed-back. The response times for the feed-back depend on the time taken for the temperatures of the various core components to change. These times are obtained from heat transfer equations relating the component temperatures to the power, coolant flow rate and the coolant inlet temperature.

The precise features of the feed-back mechanisms vary from reactor type to reactor type. In a fast reactor, the three important mechanisms of reactivity feedback under normal operating conditions are the Doppler coefficient, expansion coefficients and differential expansion coefficients. Additional mechanisms of reactivity feedback under accident conditions are by coolant voiding and fuel slumping.

3.3a Doppler coefficient of reactivity: The probability of a neutron being absorbed by or causing fission of an atom varies with the energy of the neutron and in a range of energies (0 to 20 keV), there are sharp peaks in these probabilities called resonances. There is a basic theorem (see for example, Hummel and Okrent 1970) of reactor physics that the rate of absorption of or fission by neutrons in the resonance energy range increases if the atoms are themselves in motion. Thus, when the temperature of the reactor core materials increases, the increased thermal motion of the atoms causes the reaction rates to increase. Consequently, the core reactivity changes giving rise to the Doppler coefficient of reactivity.

For fertile materials $\text{U}^{238}$ or thorium, the predominant reaction is the absorption of neutrons and hence the Doppler effect in these materials increases the absorption of neutrons and reduces the core reactivity, i.e. the feed-back is negative. On the other hand, for fissile materials both the absorption and fission rates are increased. The increased loss of neutrons by absorption is compensated by the increased production of neutrons by fission so that the net result of the Doppler effect in fissile materials is very small. The total Doppler coefficient of the reactor is the sum of the different contributions of the fissile and fertile nuclides in the fuel, with the contribution from the fertile material being dominant.
The following are the important characteristics of this coefficient:

(i) It is a prompt coefficient related to the fuel temperature, which temperature is the first to increase when the power increases.

(ii) In a small fast reactor with high enrichment, the neutrons have a short lifetime as they are quickly absorbed in the highly enriched fuel or quickly leak out of the small reactor. Hence the average neutron energy is high, close to that of the fission neutrons. The resonance energy region is about 100 times lower than the fission neutron energies and in such small reactors, there are very few resonance energy neutrons so that the Doppler effect is negligible. Thus, for example, there is a negligible Doppler coefficient in the FBTR.

(iii) In larger reactors of lower fuel enrichment and smaller neutron leakage, the neutrons have a longer life so that an appreciable number of neutrons slow down into the resonance energy region. The Doppler coefficient is thus large and is important for the stability and safety of the reactor. The coefficient is largest in oxide fuelled fast reactors which have low average neutron energies on account of the presence of two moderating atoms (oxygen) per atom of fuel.

(iv) The calculation of the Doppler coefficient in fast reactors is difficult on account of the small fraction of neutrons entering the resonance energy range from the fission neutron energies. Detailed calculations with accurate nuclear data are required (Nicholson and Fischer 1968) to establish correctly the number density of neutrons in the resonance energy range. This small fraction of neutrons contribute little to the overall neutron balance, but they are the main cause of the Doppler reactivity effect. Many years of worldwide effort have resulted in an ability to calculate this coefficient with an accuracy of about 20% (Ganesan 1979).

(v) The increase in reaction rates, as the thermal motion of the target atoms increases, tends to get saturated at higher temperatures. The Doppler coefficient hence decreases as the temperature increases. Commonly, a fitted linear decrease is used, so that \( T_{dc}/dT \) will be a constant called the Doppler constant which for the larger fast reactors varies from \(-0.001\) to \(-0.01\) according to the fuel and size of the reactor. For the carbide fuelled version of PFBR, the Doppler constant is calculated to be \(-0.0064\) for normal operating conditions and \(-0.0046\) under coolant voided conditions.

(vi) A very important series of experiments in connection with the Doppler coefficient were conducted in 1972 in the SEFOR LMFBR of USA (Harris 1973). This was a 500 litre oxide fuelled LMFBR with neutron spectrum and fuel temperatures representative of large LMFBRs. The experimental programme was designed to demonstrate that a prompt negative Doppler coefficient of sufficient magnitude makes the LMFBR an extremely stable and inherently safe reactor system. The super prompt critical transient tests involving reactivity insertion of up to 1.3 in 0.1 sec provided a convincing demonstration of the effectiveness of the Doppler coefficient in limiting the energy release in a fast reactor power excursion.

3.3b Expansion coefficients of reactivity: A rise in temperature causes the different materials in the fast reactor core to expand thereby reducing the number of atoms per unit volume. Reaction rates depend on the product of neutron density and the number density of atoms, and so all reaction rates decrease and the fraction of neutrons leaking out of the core increases. The effect on the reactivity varies according to the type of material suffering expansion.
Fast reactor safety

Fuel density reduction leads to loss of reactivity on account of reduced fuel-neutron interactions and increased neutron leakage. On account of the gap between fuel pellets and the clad inner wall, the radial fuel density reduction is in fact governed by the radial expansion of the structural steel. The fuel pellets' axial expansion is governed by the fuel temperature and is a fast acting coefficient. In small fast reactors, the Doppler coefficient is negligible and so the radial and axial fuel expansion coefficients assume importance. The axial fuel expansion coefficient for ceramic fuel changes as the fuel structure changes due to irradiation, and the study of its behaviour with burn-up of the fuel is important. Much useful information on the behaviour of this coefficient with burn-up of mixed plutonium-uranium-oxide fuel has been obtained from the operation of the French reactors RAPPIODIE and PHENIX (Berthet et al 1982; Clauzon et al 1976).

The expansion of the sodium coolant has several opposite effects on the reactivity:

(i) Reduction in the absorption of neutrons by sodium leading to a small gain in reactivity.

(ii) Reduction in the neutron moderating scattering collisions with sodium atoms. Hence, the average energy of neutrons in the reactor increases. In fast reactors, when the neutron energy increases the number of neutrons released per neutron absorption in the fuel increases, and the parasitic neutron absorption in structural and coolant materials decreases. Thus, the reactivity of the core increases.

(iii) Increase in the leakage of neutrons from the core on account of reduced sodium-neutron scattering interactions, leading to a loss of reactivity.

In small reactors like FBTR, the third effect is dominant so that the sodium expansion coefficient is negative. In larger reactors like PFBR, the third effect is small compared to the second one so that the coefficient becomes positive.

The expansion of steel material in the core causes effects similar to that of sodium, though in this case, as steel is composed of atoms much heavier than sodium, the effect due to change in neutron moderation is small and those due to change in leakage and absorption dominate. This expansion coefficient is negative in small reactors and becomes positive in large ones.

It is difficult to distinguish experimentally the differing speeds of response of the expansion coefficients associated with the fuel, structure and coolant in a fast reactor core and a total expansion coefficient with a single time constant of response is found adequate to describe the kinetic behaviour. This total expansion coefficient is always negative and has a response time of 1 to 5 sec. In FBTR, the total expansion coefficient has a value of about $-4 \times 10^{-5}/\degree C$ with a response time of about 2 sec. These expansion coefficients can be relied upon and are important for turning round slow reactivity increases and preventing system upsets leading to unsafe conditions.

3.3c Differential expansion coefficients: On account of leakage of neutrons, the neutron density and consequently the power is greater towards the centre of the reactor than towards the boundary. Further, the coolant enters at the bottom of the core, gets heated and leaves at the top. Thus, there are thermal gradients across the core and changes in these gradients can cause reactivity changes. The major effects are radial bowing of the fuel subassemblies governed by the radial temperature gradient changes; and relative movement between the control rods and core governed by the temperature difference between the hot sodium at the top in which the control rod shafts dip, and the cooler sodium at the bottom around the core support structures.
Inward bowing of the fuel subassemblies leads to an increase of density of core materials in the core centre, where the neutron density is higher. This leads to an increase in reactivity. Such a positive bowing coefficient was present in the early fast reactor EBR-I and led to a core meltdown accident (Thompson and Beckerly 1964) during special experiments for measuring the feed-back coefficients. Modern designs for LMFBR's incorporate spacer pads at suitable levels between fuel subassemblies to make the bowing coefficients zero or slightly negative. Thus FBTR has a slightly negative bowing coefficient and PFBR would be also designed to have a negative one.

The reactivity coefficient associated with control rod shafts expansion is a fairly large negative one but acts after a time lag. In the case of FBTR, it is expected to have a value of \(-2.6 \times 10^{-5}/^{\circ}{\text{C}}\) with a response time of 25 sec.

3.3d Sodium void coefficient: Exactly similar arguments as for the reactivity change during the expansion of sodium show that there will be a reactivity change if the sodium coolant is accidentally voided from the core. This is called the sodium void coefficient and is negative for small cores like FBTR and positive (+3 dollars) for large cores like PFBR. Obviously, there are strong safety implications associated with the positive coolant voiding coefficient, whose value can attain several dollars for a large power reactor. Comprehensive theoretical and experimental investigations for the prediction of this coefficient have been made leading to a satisfactory understanding (Küsters and Ganesan 1978). Currently, the successful designs to have a lower sodium void coefficient are the so called heterogeneous cores in which internal blanket material zones are placed within the core (see for example, Inoue et al 1982). In this case, voiding leads to neutron leakage into the internal blanket zones thus reducing reactivity. However, no heterogeneous core fast reactor has been operated. The alternate approach has been to maintain a positive voiding coefficient and ensure by proper safeguards and design that the sodium cannot be voided suddenly. This has been found particularly easy for pool type reactors where the possibility of loss of sodium by pipe break is not possible and where due to thermal inertia, there is considerable time lag for the sodium to boil when there is a loss of flow. Thus, for example, both for the 250 MWe PHENIX and the 1200 MWe SUPERPHENIX reactors, a loss of electric supply to the primary and secondary pumps followed by a failure to shut down the reactor does not result in sodium boiling (Gouriou et al 1982).

3.3e Fuel slumping reactivity gain: Unlike thermal reactors, in a fast reactor the spacing between the fuel pins is governed by heat removal considerations and not from reactivity considerations. Compaction of fuel tends to reduce the neutron leakage and increase the reactivity. Hence, fuel melting and slumping is the most serious accident in a fast reactor. Traditionally, the worst rates of reactivity addition were calculated from a gravitational collapse of the core assuming gross core melting due to loss of flow of coolant. The total reactivity addition can be several tens of dollars and the rate of reactivity addition can be of the order of a hundred dollars per second. The smaller the reactor, the greater is the reactivity gain due to slumping. It is to be noted that the values for the slumping reactivity gain and the rate of reactivity addition given above are extreme theoretical values and realistic modelling of accident sequences does not give such values. However, it was these high values which were the early causes for concern and gave the impetus for extensive fast reactor safety research.
3.4 Engineering features

There are many engineering design features which enhance the inherent safety of the LMFBR. An important feature is that the sodium system is a low pressure system and the maximum sodium temperature (~ 550°C) is far below the sodium boiling point. Further, sodium has extremely good heat transfer and natural convection properties which ensure good decay heat removal even under accident conditions. General design features enhancing the safety are (Graham 1971):

(i) use of special engineering design codes for fast reactors with appropriate safety factors,
(ii) limited rates of control rod withdrawal,
(iii) use of diverse and redundant shutdown systems,
(iv) provision of flywheel action in main sodium pumps to maintain flow under loss of power,
(v) provision of alternate grid supplies,
(vi) use of multiple radial inlets to subassemblies to avoid flow blockages,
(vii) use of wrapped and restrained piping to avoid pipe breakage; use of safety guard tanks around components like reactor vessel, iHX, pumps, etc.
(viii) use of double walled piping, check valves, hydraulic diodes, reserve coolant volumes, etc.

The design features providing inherent safety in FBTR are diverse. The coolant entry to subassemblies is via multiple (twelve) openings which cannot get blocked simultaneously—even with 50% blockage the flow reduction is only 10%. Hold-down springs are provided at the feet of subassemblies even though the sodium fluid dynamic forces are insufficient to lift a subassembly. In addition to flywheels which are provided on pumps to prevent abrupt flow stoppages, diverse and independent power supplies feed the pumps including emergency batteries and diesel generators. The general safety features which would be incorporated in PFBR include a pool type design using a large double walled vessel to house the reactor. Multiple inlets for sodium entry into subassemblies are to be provided. Emergency power supplies, flywheels on pumps and two independent shutdown systems will be provided.

4. The plant protection system

4.1 Elements of the plant protection system

Despite the assurance offered by careful design, construction and operation, the second and most important level of safety is the pps, which can handle a wide range of conceivable abnormal incidents and malfunctions and safely shut down the reactor. Basically the system consists of:

(i) a wide variety of instruments to monitor the plant operating parameters and characteristics,
(ii) assured shutdown systems triggered by signals from the monitors,
(iii) decay heat removal systems.

Core damage may occur if the shutdown systems or the decay heat removal systems fail on demand. Hence proper operation of the pps is very important. This is assured by means of redundancy, diversity and fail-safe operation of the components of the pps.
Redundancy refers to the use of two or more similar components in parallel by means of which it is assured that the pps does not fail when needed, and at the same time, false plant shut-downs are not caused by instrument errors. Diversity in shutdown systems, electric power systems, etc. is provided to prevent common mode failure, i.e. a number of failures resulting from a single cause. By fail-safe operation is meant that failure of a component of the pps results in shutdown of the plant rather than continued unprotected operation. Recent designs, like the SNR 300 and SUPERPHENIX, provide for two independent plant protection systems (Birkhofer et al 1982; Berlin et al 1982).

The pps often includes a process computer which monitors and processes the information of all signals relevant to the safety of the plant. In FBTR, this is performed by the Central Data Processing System (a pair of ECIL TDC-316 computers) which can also initiate alarms and safety actions.

4.2 Plant monitoring systems

The most important monitoring systems in a fast reactor plant are for neutron and sodium temperature and flow monitoring; and clad rupture detection. Other monitoring systems are for sodium levels and sodium leaks and aerosol monitoring; and steam generator leak detection.

4.2a Neutron monitoring: Neutron detectors and associated electronics are required in order to monitor and control the power level as well as to provide signals for safety action. The neutron monitoring system provides signals for alarms, power reduction or reactor shutdown when the neutron density level is too high or the rate of increase of neutron density is too high. In addition, most LMFBRs contain reactivity meters which use measured neutron density as a function of time to derive the reactivity changes, which can also signal for safety action if found excessive. The neutron monitoring system is triplicated and safety action is taken when any two of the three monitors call for it.

On account of high gamma radiation fields, all neutron detectors should discriminate against gammas. Both for FBTR and PFBR, the neutron detectors considered are BF$_3$ counters, boron coated counters, fission counters and boron coated compensated ion chambers. Location of the detectors outside the primary vessel is desirable from considerations of temperature and radiation. In FBTR, the detectors are located at the start of the radial concrete shielding outside the reactor vessel. For PFBR, location of detectors under the main vessel is under consideration. Further, for a pool type reactor like PFBR, adequate neutron levels are not obtained outside the main vessel for shutdown monitoring and use of in-vessel detectors is to be considered.

4.2b Sodium temperature and flow monitoring: Monitoring of sodium temperatures and early detection of sodium boiling is very important for core safety especially with positive coolant voiding coefficients. Most commonly, thermocouples (Chromel-Alumel or Ni-Cr) and Resistance Temperature Detectors (RTD) are used. In FBTR, each fuel subassembly is monitored by 2 thermocouples and a similar system may be considered for PFBR.

Sodium flows are measured both by Venturi flow meters and magnetic flow meters. While the former are very accurate, their response time is too slow for use in the pps, which purpose is served by the latter.
Trip signals are actuated by high sodium temperature, low flows, mismatch of power to flow values, loss of electric power supply, etc.

The onset of sodium boiling is also sought to be detected in modern LMFBR's by the methods of neutron noise and acoustic noise detection. Bubbles formed when sodium boils give rise to both reactivity fluctuations (see the discussion in §3.3b) as well as to pressure variations in the metal. These fluctuations can be recorded by neutron monitors or by pressure transducers and the onset of boiling detected.

4.2c Clad rupture detection: Detection, localisation and removal of failed fuel with ruptured clad is required to prevent excessive contamination of the primary coolant and to prevent flow blockages that could arise on account of gradual fuel swelling when sodium contacts the fuel. Both in FBTR and in PFBR, the methods of failed fuel detection are to be by monitoring of the cover gas fission product activity and by monitoring of delayed neutrons in the primary coolant. Excessive activities from these monitors trigger a plant shutdown.

The cover gas fission product activity is due to gaseous fission products like Xe and Kr escaping from ruptured fuel pins. These gases require time to reach the cover gas plenum and the response time for clad rupture detection by cover gas monitoring is of the order of minutes. On the other hand, fission products like I\(^{137}\) and Br\(^{87}\) which are delayed neutron precursors get dissolved in the sodium upon clad rupture and the emitted neutrons are detected in a bypass loop of the primary circuit. The response time of this system is a function of the coolant transport time and is in the range of 20–30 sec. Identification of the failed fuel subassembly is an important problem and one of the methods followed in some reactors is to sample the sodium from each subassembly by selector valves and monitor the delayed neutrons.

4.3 Reactor shutdown systems

Besides the function of reactivity or power adjustment, the control rods must be able to shut down the reactor under any foreseen conditions. It is important to assess the mechanisms of accidental reactivity insertions and ensure that the shutdown system can compensate for these. Such mechanisms include melting of a fuel subassembly, sodium voiding, inadvertent withdrawal of a control rod, introduction of moderator, sudden flow increases, etc.

In FBTR, the reactor control is by six enriched boron carbide control rods. The power to drive the rods is supplied independently to each of the six control rod drive mechanisms and in the event of power failure or a rapid shutdown trip action the drive mechanisms and rods fall under gravity in about 400 m sec to shut down the reactor. The electrical connections are such that only one rod can be withdrawn at a time at a rate of 1 mm/sec which limits the reactivity addition rate to a low 6 \times 10^5/\text{sec}.

As in modern LMFBR's, PFBR will have two independent shutdown systems. The primary shutdown system is expected to consist of 9 control rods to be used for normal operations as well as for rapid shutdown. In addition, a completely independent secondary shutdown system of 3 control rods based on a different design principle would be able to override abnormal reactivity additions and shut off the reactor even if all the primary rods fail to act. Their design would be such that they can enter even distorted cores, and would automatically operate on sodium overheating due to Curie point demagnetisation of magnets. Alternate design principles for the secondary system under consideration are:
(i) Articulated links with spherical couplings (Pignatelli et al 1982).
(ii) Hydraulically suspended absorber balls (Specht et al 1976).

Fuel loading and unloading in an LMFBR is done with all control rods fully inserted and disconnected from their drive mechanisms. It is important to ensure that the core is not made critical during refuelling. Hence, accurate measurement of the degree of subcriticality of the shutdown core and the reactivity worths of the loaded subassemblies is important.

4.4 Decay heat removal

The decay heat is initially about 6 to 7% of the reactor power and gradually falls with time becoming around 1% after an hour. An important part of the pps is the provision of adequate decay heat cooling under shutdown conditions for as long as necessary (up to a month or more) even under conditions of coolant system failures.

Decay heat removal in fast power reactors is ensured by a diversity of means (Lauret et al 1982; Agarwal and Guppy 1981):
(i) multiplicity of primary heat transport loops ensuring normal decay heat removal by the operation of a single loop alone,
(ii) redundancy and diversity in power supplies to pumps,
(iii) forced circulation in the immediate period following shutdown (when decay heat generation rates are high) by design provisions, such as flywheels for gradual coastdown of pumps,
(iv) continued forced circulation by low speed operation of pumps by means of pony motors backed by alternate power supplies/emergency diesel generator sets/batteries,
(v) provision of independent back-up decay heat removal systems,
(vi) design of normal as well as back-up heat removal systems, such that setting-up of good heat removal by natural convection is facilitated.

In pool type LMFBR's, the large thermal capacity of the sodium bulk acts as a heat sink and allows a long time interval (at least several hours) for alternate measures to be taken before the sodium boils. The possibility of complete loss of sodium from the primary heat removal circuits is also not present in such reactors.

For FBTR, decay heat removal is ensured by a variety of measures starting with the operation control that the reactor will be operated only with both heat transport loops. Thus one loop will always be available to remove decay heat under upset conditions. Further, emergency batteries and diesel generators provide power for the pumps on grid non-availability. Natural convection is found to be able to remove about 300 kW of heat with none of the pumps operational. In case the primary circuit is not available due to sodium leak, decay heat removal is by the emergency cooling system, where cooling nitrogen is circulated in the double envelope of the reactor vessel. Further, flooding tanks full of sodium have been provided to dump sodium into the reactor vessel and maintain the sodium level in case of a primary circuit leak. Finally, the biological shield cooling system can also remove decay heat. These provisions are considered adequate to provide decay heat cooling for any upset conditions.

In the case of PFBR, there are to be four independent loops and normal decay heat removal is ensured when any one of the loops is operational. Physical separation of the loops will be provided to avoid common failures. To provide forced cooling following shutdown, design provisions are made: flywheel inertia for appropriate pumps coast-down, and pony motors backed by alternate power supplies/diesel generators/batteries.
Setting-up of natural convection for heat removal would be ensured by suitable elevations of core, IHX and terminal sodium/air heat exchangers. It is estimated that two loops in natural convection could provide the required decay heat removal capacity. To provide cooling in case of common mode failure resulting in non-availability of all normal loops, an independent back-up decay heat removal system would be provided.

5. Principles of containment

5.1 Containment barriers

An operating commercial-sized fast breeder reactor would have an inventory of a few thousand kilograms of plutonium and a few hundred kilograms of fission products. The total radioactivity is in millions of curies. As far as environmental effects are concerned, the main nuclides of importance are the plutonium isotopes, Am$^{241}$, Cm$^{244}$, tritium, Kr$^{85}$, Sr$^{90}$, I$^{129}$, I$^{131}$, Xe$^{133}$, Cs$^{134}$, Cs$^{137}$ and the activation product C$^{14}$. The potential for escape varies according to whether the radioactive material is gaseous (like Xe or Kr), volatile (like I, Br, Cs, Rb) or non-volatile. The release of radioactive material to the environment during normal operation and under accident conditions is prevented by three physical barriers.

The first barrier is the cladded fuel element. Non-volatile fission products account for some 98% of the generated radioactivity and are retained in the solid fuel material. The fuel material itself, as well as the volatile and gaseous fission products are contained by a gas-tight metallic (steel) clad.

The second barrier is the gas-tight primary coolant boundary, which also retains, along with the radio-active primary sodium and cover gas, the radioactive material that may escape from failed fuel elements. This barrier is constituted by a double walled reactor vessel, bolted down top plugs, etc. Monitoring of coolant and cover gas activity enables shutdown of the plant and clean-up action to be taken before the radioactivity in the primary circuit reaches unacceptable levels.

The reactor containment building, usually of reinforced concrete and steel, provides the third barrier to the escape of radioactivity. Different types of containment buildings have been considered for fast reactors (Graham 1971; Seeman and Armstrong 1978). A single containment building with a low leak rate to the outside has been used in small fast test reactors like EBR-II, RAPSODIE, FFTF, JOYO, FBTR, etc. Double containment has been used for the older reactors EFFBR (60 MWe), SEFOR as well as for recent large reactors like SUPERPHENIX-1 (1200 MWe) and SNR-300. These have within the containment building, a steel containment dome with inert gas around the reactor. The 250 MWe prototype fast reactors PFR and PHENIX had been provided with leak resistant reactor buildings without regular containment building features.

The design of these successive radioactive barriers is such that the radioactive release under normal operating conditions is acceptably low as per the radiological guidelines for the chosen site and the radioactivity is adequately contained under accident conditions.

In the case of FBTR, the rotating plugs at the top of the reactor vessel are fitted with a plug restraint mechanism with hold-down bolts. These have been designed to withstand and limit the radioactivity release into the reactor containment building under certain design basis accident scenarios. The strength of the containment building and its
leakage rate are designed such that this radioactivity release accompanied by sodium fires is easily contained with no hazard to the public. Similarly, in PFBR, the double-walled reactor vessels enclosed in concrete vaults with top closures are to be designed to withstand design basis accidents without release of radioactivity.

5.2 Energy release calculations

In order to ensure the adequacy of the engineered safety features and the containment of radioactivity under all possible circumstances by the reactor vessel and containment building, it is usual to study extremely unlikely but assumed credible accident sequences and then estimate the energy release and other consequences. These hypothetical accidents assume complete failure of the plant protection system and are called design basis accidents. Their precise features vary from reactor type to reactor type. In fast reactors, the design basis accident is taken as fuel compaction resulting from meltdown of fuel leading to what is called a Hypothetical Core Disassembly Accident (HCDA) whose mechanical work has to be assessed to adequately design the engineered safeguards.

The classical model for the energy release in an HCDA is that due to Bethe and Tait (1956) which has been improved upon by later workers (e.g. Nicholson 1964). The model assumes an initial large reactivity insertion rate from fuel slumping which then leads to a power excursion and energy build-up in a molten core which is limited but not terminated by the Doppler coefficient. Pressures then build up in the fluid core which expands and terminates the excursion.

Important information required is the equation of state of core materials upto high temperatures (≈ 7000°C) which are not always accurately available. The rate of reactivity addition is a second important parameter on which very much depends the value of the final energy release. This rate depends on the kind of accident initiation. The Doppler coefficient is a third important parameter governing the energy release. Calculations indicate (Wilson 1977) that a Doppler coefficient with \(-\frac{\Delta k}{\Delta T}\) greater than 0.002 is sufficient to effectively reduce the energy build-up by a factor 100. Lastly, the initial power and temperature as well as the amount of void space that the fuel vapour has to fill before the disassembly can start, are important initial conditions governing the final energy release.

Computer codes like VENUS-II (Jackson and Nicholson 1972) calculate the heat energy developed accurately, subject to the uncertainties in the initial conditions. However, calculation of the conversion of the heat energy to mechanical energy or more precisely the forces and pressures which can rupture the containment barriers is more problematic. Research on fuel coolant interaction (FCI) and shock structure interactions are of relevance here. Generally, isentropic expansion of the vapour of all core materials is taken as the mechanism of conversion to mechanical work, and recent calculations with the comprehensive computer code SIMMER-II (Smith 1978) show that this approach is safe and conservative.

As noted earlier, the energy release in an HCDA is dependent on the initial conditions which in turn depend on the details of the accident initiation and the accident sequence. Hence, a detailed mechanistic modelling of the accident sequence is necessary for realistic estimates of the energy release. This has given rise to a vast amount of research in this field.

For a mechanistic analysis of design basis accidents, various possible accident
initiators have been considered (Graham 1971) and the accident sequences usually studied are the unprotected Loss of Flow Accident (LOFA) and the unprotected Transient Over Power Accident (TOPA) resulting from a postulated failure of the PPS along with either sodium pumps failure or accidental control rod withdrawal. Large, sophisticated computer codes like SAS4A (Jackson and Stevenson 1981) have been created to trace these accident sequences. The in-pile safety related experimentation in the TREAT and the CABRI facilities have provided very important data for the successful modelling of the design basis accidents (see for example, Spencer et al 1976; Dadillon et al 1982).

It is important to note here that in the last few years, there is a re-evaluation of the role of HCDA from the probabilistic viewpoint. The HCDA is found to have no initiator in a modern LMFBR (Fauske 1981) and should not influence the plant design. Thus, detailed mechanistic considerations beyond early termination is to be considered redundant.

In the case of FBTR as well as PFBR, design basis accidents have been studied in detail for various fuel types and core configurations using computer codes PREDIS (Singh et al 1981) and VENUS-II. The results of these analyses have provided the basis for the containment design.

6. Conclusion

A review has been made of the principles of fast reactor safety and the different factors contributing to the overall safety of a modern LMFBR have been highlighted. The important conclusion is that the design, construction and operation of an inherently safe, well protected LMFBR pose no insurmountable problems, and much of the uncertainties surrounding fast reactor safety issues have been resolved by contemporary research.

Acknowledgements

The above review is based on detailed studies made by scientists and engineers at RRC for the FBTR and PFBR projects. Thanks are due to Dr O P Singh for several useful discussions in connection with §5.2.

References

Bethe H A and Tait J H 1956 Report UKAE5-RHM (56)/113


Ganesan S 1979 Studies on sodium void and Doppler reactivity effects in fast power reactors Ph. D. Thesis, Bombay University

Glasstone S and Edlund M C 1952 The elements of nuclear reactor theory (New York: Van Nostrand) p. 82

Glasstone S and Edlund M C 1952 The elements of nuclear reactor theory (New York: Van Nostrand)


Hanson J E 1984 Trans. Am. Nucl. Soc. 46 331

Harris R A 1973 Report Hedl-TME-73-42

Hummel H H and Okrent D 1970 Reactivity coefficients in large fast power reactors (Hinsdale: American Nuclear Society)


Keepin G R 1965 Physics of nuclear kinetics (Reading Massachusetts: Addison-Wesley)


Lauret, Minguet, Cuttica and Barracco 1982 Proc. LMFBR safety topical meeting (Lyon: European Nuclear Society) p. II-551


Oakes L C 1968 Nucl. Saf. 9 363


Smith L L 1978 Report LA-7515-M


Wilson R 1977 Rev. Mod. Phys. 49 893
Engineering development and safety back-up for nuclear power programme

P R DASTIDAR and A KAKODKAR
Reactor Group, Bhabha Atomic Research Centre, Bombay 400085, India

Abstract. This article reviews the engineering development and safety aspects that are relevant to the nuclear power programme being pursued in the country. Some of the important aspects have been discussed in detail bringing out the current status and also the directions for further work.

Keywords. Nuclear power programme; engineering development; safety back-up; pressurised heavy water reactor; emergency core cooling system; thermosyphon; coolant tubes

PACS No. 28.45; 28.40

1. Introduction

It is now well known that nuclear power is the only source of energy for bulk power generation which can meet our future requirements. Our resources of fossil fuel which form the mainstay for bulk electricity production at present would tend to get depleted with time and it is necessary to develop the alternate nuclear source to a level where it can take over from the fossil fuel sources while maintaining the growth of electricity generation necessary to sustain the national development. The three-stage nuclear power programme envisaged by the Department of Atomic Energy is aimed at achieving this objective. As of now the emphasis has been mainly towards mastering of the nuclear technology and to demonstrate that nuclear power stations can be built within the country with nuclear electricity available at a cost which is competitive with the conventional electricity. Time has now come when this effort needs to be accelerated to enhance the installed nuclear capacity. The nuclear power programme of 10,000 MWe by the year 2,000 outlined by Dr Ramanna is our immediate objective. In this phase, not only one would increase contribution of nuclear electricity but also enable production of sizable quantities of plutonium which, with the available uranium resources, can sustain a much larger fast breeder programme. It is envisaged that through the use of fast breeder reactors an installed capacity to the tune of 350,000 MWe can be realised with available uranium resources. Subsequent to this, further generation of nuclear electricity would be sustained by large reserves of thorium resources.

Realisation of the objective of 10,000 MWe capacity by the year 2,000 is envisaged to be achieved by installing a number of 235 MWe and 500 MWe power units of PHWR type. In order to realise installation of such a large number of units in a small time, many aspects such as availability of finance, manpower, industrial capacity, organisational setup, safety regulation etc have to be looked into in detail. The purpose of this
article is to review the engineering development and safety back-up that would be necessary to support the proposed power programme.

2. The PHWR system

Our power programme is based on pressurised heavy water type reactors. We are now at a stage when for the 235 MWe units we have gone through a full cycle of design, construction, commissioning and operation. Although the experience with operation of RAPS-I unit, which has been the first in the PHWR series of reactors, did reveal some inadequacies, the performance of RAPS-2 and MAPS-1 has been highly encouraging. Even for RAPS-1 the lessons have been learnt and it should be possible to bring this reactor back on stream quite soon. It is worth mentioning at this stage (Strachan and Brown 1984) that similar stations in Canada have shown excellent performance. As a matter of fact, in terms of lifetime capacity factors up to end of 1981, seven out of 8 Candu units, find a place in the first ten positions of the world’s most reliable operating plants of capacity greater than 500 MWe. The eighth one has 18th position. The design of reactor units beyond Narora considers new features that have been incorporated in Canadian design as well as our own experiences. There is thus no doubt that the PHWR units that are being added would show even better performance. There is however, always a need to look for further improvements on a continuing basis. For the 500 MWe units which are now under design this need is even greater.

For realisation of improved performance as well as safety characteristics of the reactor units it is necessary to evolve new systems and generate adequate data to understand and predict their behaviour under a variety of conditions that could be expected during service. We would be reviewing some of these areas in this article.

Figure 1 shows a schematic of pressurised heavy water reactor power plant. Reactor core is housed in a cylindrical calandria which contains heavy water moderator and reflector at low pressure. A large number of (306 in 235 MWe units and 392 in 500 MWe units) fuel channels pass through the calandria. These channels contain fuel and also flowing coolant at high pressure thus permitting heat removal from the core. The primary coolant circulating pumps circulate the coolant through the calandria. These channels contain fuel and also flowing coolant at high pressure thus permitting heat removal from the core. The primary coolant circulating pumps circulate the coolant through the core and the steam generators. The heat extracted from the core by the coolant is transferred in the steam generators to raise steam which drives the turbo generator, thus producing electricity. A number of other important systems also have to function. A moderator

![Figure 1. Pressurised heavy water reactor power plant.](image-url)
heat removal system, on-power fuel handling system and reactor regulating system are some of the important systems. In addition, a number of safety systems are provided to protect the whole system, plant personnel and members of public. A view showing more details of the reactor is shown in figure 2.

The PHWR system has been an ideal choice for our power programme. It is based on natural uranium and heavy water. Technology for both has now been developed and demonstrated on industrial scale. Manufacturing technology for various equipment has also been successfully mastered. The system has a number of safety features not available in other reactor types. These are: (i) The actuators for reactor protection act in low pressure environment unlike the popular PWR or BWR reactors. Thus PHWRs are immune to accidents that can be caused by ejection of neutron absorber from the core. (ii) The excess reactivity of the core is not required to be high since on-power refuelling is available to compensate for the fuel burn-up at any time. (iii) Cold moderator is
always available in the core. In extreme situations this heat sink comes into play in removing heat from the fuel, thus limiting maximum temperature to a lower value. (iv) High pressure reactor pressure vessel is eliminated.

In contrast the channel type design of PHWR presents some safety considerations that are different from those of other designs. These are: (i) provision of adequate cooling to all channels under accident situations (ii) effects of the possibility of fuel coming in contact with the pressure boundary, (iii) introduction of positive reactivity in an unlikely loss of coolant accident event.

The safety criteria to be used for assessment of the system therefore have to consider the above characteristics which are distinct and specific to PHWR. Need for additions or improvements has also to be judged based on these characteristics.

3. Evolution of safety approach

The safety approach has to encompass a number of situations. During normal operation the permissible maximum individual dose limits as well as population dose limits have to be satisfied. In this category one would also include all off normal situations of rather frequent occurrence. Next we have to take care of various accident situations. Clearly both failure probability and consequences have to be considered. The known difficulties with regard to certain accident situations like stagnation breaks, anticipated transients without scram and common cause effects of earthquake have led to the philosophy of two group concepts with each function being provided for by independent systems from each group (see figure 3). With this philosophy it is possible to categorise accident situations in two categories (Atchison et al 1983). One category consisting of what are called serious process failures (single failure) and the other consisting of events which are dual failures i.e. failure of the process equipment along with failure of a safety system. In the later case accident consequence is mitigated by the second safety system.

The next step is to assign maximum frequencies and the reference dose limits for each of the above two categories and to ensure that possibility of triple failure i.e. failure of process as well as safety systems in both groups is made acceptably small i.e. around $10^{-7}$/yr.

Talking in engineering terms, the following guidelines evolve. (i) The design, construction and operation of all components, systems and structures essential to

<table>
<thead>
<tr>
<th>Main control room</th>
<th>Aux. control room</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group - I</strong></td>
<td><strong>Group - II</strong></td>
</tr>
<tr>
<td>Process system</td>
<td>Shut down system</td>
</tr>
<tr>
<td>- Reactor regulation</td>
<td>Containment</td>
</tr>
<tr>
<td>- All process systems</td>
<td>Aux. moderator cooling</td>
</tr>
<tr>
<td>except aux. moderator cooling</td>
<td>Emergency water and power supply</td>
</tr>
<tr>
<td>Safety system</td>
<td>ECCS</td>
</tr>
<tr>
<td>- Shut down system-I</td>
<td></td>
</tr>
<tr>
<td>- ECCS</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Two-group concept for safety system.
safety of the reactor will follow the best applicable codes, standards or practices and confirmed by an independent audit. (ii) The quality and nature of the essential process equipment will be such that the total of all serious process failures should not exceed a specified number. A serious process failure in this context is a failure of process system or equipment that, in the absence of safety system action, could lead to fuel failure or the release of radioactive material to the environment. (iii) The safety systems belonging to two groups will be physically and functionally separate from the process system and from each other and (iv) each safety system will be readily testable and will be tested at a frequency that demonstrates its unavailability to be less than a specified number.

Consideration of various dual failures also leads to definition of performance requirements for safety systems. For example, a loss of coolant accident (LOCA) plus failure of emergency core cooling system will lead to release of fission products from the fuel (the “source term”) that must be accommodated by the containment system. Similarly, a LOCA with impaired containment system sets the effectiveness required of the ECCS. Some of the concerns which are not satisfactorily answered following the single/dual failure approach are:

(a) inability to consider large variation in frequency and consequences of different single/dual failures. To overcome this difficulty the event frequency vs permissible dose limit acceptance curve is divided into a larger number of steps;
(b) problems relating to safety system support features such as electrical, air or water supplies whose failure could result in failure of both process as well as safety system;
(c) need to ensure continued operation of safety system after an accident and
(d) need to include consideration of accidents like earthquake which could damage both process and safety system.

A more comprehensive approach to safety evaluation is thus needed. This leads to the next stage of evolution—which is based on use of safety design matrix (SDM). It is essentially a systematic record of visual inspection by the analyst of all selected events of potential safety concern. Extensive use of fault tree analysis and event sequence diagram enables identification of possible causes for the event and also the various consequences. SDM enables better understanding of the system behaviour and has the potential to identify desirable design modifications. This approach can be extended to predict the risk posed by any postulated event sequence.

Progress is being made towards application of probabilistic risk assessment (PRA) techniques. With development of this technique and also appropriate data base it appears that PRA would have a strong influence on licensing decisions in future.

4. Engineering development

Engineering development is a process of continuous evolution. The availability of new technologies, evolution of new safety criteria, experience with design, construction, commissioning and operation of power reactors, availability of new information with regard to design techniques, material behaviour, equipment reliability, understanding of accident behaviour etc all have an impact on the new engineering development effort. As mentioned earlier the PHWR system is expected to perform well both in terms of safety and in terms of its performance as a power unit. However, there is always a scope
to improve the system through new development. Let us now review some of the important areas in greater detail.

5. Emergency core cooling system

From the point of view of performance of emergency core cooling system the distinctive features of PHWR cooling system in contrast to the more popular PWR systems should be recognised. There are: (i) The length of large diameter piping is quite small in PHWRS. In contrast the small diameter piping is far more extensive in PHWR. The PHWRS have thus a larger probability for small break LOCA. (ii) The coolant tubes which contain and also support the fuel are horizontal. They can deform into different modes depending on the pressure temperature history. (iii) The primary heat transport system is arranged in a figure of eight configuration (figure 4). Certain combinations of size and location of break can cause stagnation conditions in the core. (iv) Voidage of core introduces positive reactivity.

It is thus clear that the system behaviour in case of LOCA in a PHWR is likely to be quite different from the pressure vessel type reactors for which most of the published literature is relevant. Thus while the design of ECCS goes on based on the best available calculational procedures, an extensive experimental programme to study various effects

![Figure 4. Figure of eight configuration.](image-url)
that would be manifested during various stages following LOCA is also being pursued. This work can be divided into two broad areas. One relating to the blowdown phase and the second relating to emergency core cooling phase.

In the blowdown phase the main feature that distinguishes the blowdown in PHWR is the fact that the fuel channels in PHWR are horizontal. There is thus need to understand through experiments the various phenomena that are involved in blowdown from horizontal channels. Apart from rigging up the experimental set-up, this also involves development of techniques for measurement of blowdown rates with a variety of break sizes. In parallel to the experimental programme, analytical studies aimed at development of a computer program to calculate transient two-phase flow and heat transfer during blowdown from horizontal pipes starting with subcooled condition and with heat addition need to be pursued (Venkat Raj et al 1983).

In case of PHWRs this analysis needs to be extended to analyse the effects of failure of ECCS (Hancox 1981). In such an event the build-up of temperature would lead to deformation of coolant tube which could sag into contact with the calandria tube if the coolant pressure is already low. Alternatively if the pressure has not sufficiently dropped down, the coolant tube can balloon into contact with the calandria tube. The two possibilities are pictorially depicted in figure 5.

Once the contact with the calandria tube is established heat can get dissipated to moderator and the maximum fuel temperatures get decided by the resistances in this heat transfer path.

In the ECCS phase, the specific areas of concern to PHWRs are: (i) problems caused by the injection of cold emergency coolant in voided header such as condensation induced

![Figure 5. Schematic diagram of the stages of fuel channel heat up and deformation.](image-url)
flow oscillations; (ii) flow distribution between parallel channels under ECCS conditions. (iii) Rewetting of hot horizontal channels.

Here again extensive experimental data are required to enable analysis of conditions in the core during LOCA. Depending upon the temperature of the fuel in various channels, it is possible that a larger share of the flow injected in the header may get preferentially diverted to the cooler channels contrary to the actual requirements. Model studies are therefore required to investigate this phenomenon and to ensure that all channels receive at least the required quantity of coolant flow.

Considerable work has been done to understand the rewetting phenomenon in horizontal rectangular as well as annular ducts (Venkat Raj 1984). A typical experimentally determined rewetting transient starting with an initial temperature of 500°C is shown in figure 6. Comparison of traces CB, CM and CT indicates an inclined rewetting front. Significant stratification, flow chugging and oscillations in pressure drop have also been observed. A comparison of experimental data on rewetting velocity with theoretical prediction for an annular horizontal duct is given in figure 7. The data collected during such tests can be used to work out a more realistic prediction of the

---

**Figure 6.** Rewetting transients (initial temperature — 500°C).

**Figure 7.** Comparison of predicted rewetting velocity for Ants-2 with experimental data.
integrated system behaviour following a LOCA and the performance of the ECCS. Such a behaviour can also be studied on specially designed experimental set-ups.

6. Thermosyphon

The current designs of PHWRs depend on the thermosyphon for maintaining the circulation through the core under conditions when the main circulating pumps are not running and the system pressure and temperature do not permit bringing in the standby heat removal system (pumps and heat exchangers). While for the new reactors this situation can be avoided by incorporating a standby heat removal system capable of working at full pressure and temperature, a study of thermosyphon phenomenon would be useful both for existing reactor systems as well as future reactor systems. Figure 8 shows a plot of flow (in terms of temperature difference) vs time determined on an experimental set-up simulating the PHWR primary system layout. This plot indicates both initial establishment of flow as well as stability of flow under single phase flow conditions.

Experimentally determined data for channel pressure drop over the full range of Reynolds numbers are shown in figure 9. Similar data need to be generated for the coolant channel of 500 MWe PHWR. Such data are essential for evaluation of safety in a number of situations when flow is less than normal flow.

Study of thermosyphon in the PHWR system under two phase flow conditions also needs to be pursued.

7. Zr-2½-% Nb coolant tubes

Our PHWRs so far use zircaloy-2 as the material for the coolant tubes. This material was also used by Canadians in their earlier reactors. With the development of Zr-2½-% Nb material which has better strength and creep resistance Canadians switched over to this new material starting from pickering-3 unit. The primary objectives have been to extend the creep life of coolant channels and effect greater neutron economy.

A number of problems have been faced by Canadians in the use of Zr-Nb tubes. Most of them have been caused by delayed hydride cracking caused by faulty rolling
Figure 9. Fuel channel pressure drop data.
procedure. Although all such tubes have been subsequently stress-relieved, it is
anticipated that some initial cracking might have been caused during the operating
period between faulty rolling leading to high residual stress and the actual stress
relieving operation. The cracks initiated during this period would progressively grow
and tube failures on this account are expected even in future for initial reactors tubed
with Zr-Nb tubes. The mechanism of delayed hydrogen cracking has now been
understood well and the new rolling procedures have been developed which minimise
the residual stress thereby eliminating delayed hydride cracking and also making the
stress relief operation unnecessary.

A new mechanism has come to light consequent to failure of zircaloy-2 pressure tube
in Pickering-2. Displacement of garter spring support between the coolant tube and the
calandria tube led to sagging of coolant tube in contact with the calandria tube. The
cold spot so developed on the outside of coolant tube developed blisters of hydride
almost along the line of contact. A fairly long partial thickness crack got generated
through cracks in these blisters. When the crack exceeded the critical dimension there
was a sudden unstable propagation of crack leading to gross failure of the coolant tube.
This event has led to violation of the belief held till then that any failure in the coolant
tube would be of leak before break type. On the positive side of this event were the facts
that the calandria tube did not rupture in this event and also that the reactor was shut
down in an orderly manner demonstrating that such events can be safely handled. A
large amount of literature is being published indicating that the hydrogen pick-up rates
in Zr.Nb tubes are much less as compared to zircaloy-2 tubes.

In view of the serious safety implications this matter has been very carefully reviewed.
A decision has been taken to adopt Zr-Nb coolant tubes for future 235 MWe units as
well as the 500 MWe units. Needless to say that additional features like increase in
number of garter springs, use of gartersprings immune to unwanted displacement etc
have also to be looked into at the same time.

An extensive programme is already underway to develop and evaluate Zr.Nb coolant
tube production at the Nuclear Fuel Complex, Hyderabad, and to develop a coolant
channel based on this new material. Clearly the most crucial activity in this work is the
development of rolled joint which would not pose any delayed hydride cracking
problem.

8. Structural safety

Failure of piping and other important structures could have significant safety
implication. For example LOCA is caused by a sudden failure of piping or some other
component in the primary heat transport system. Development of an identifiable leak
well before sudden fracture is a preferable situation as it generates an early warning to
enable corrective action. Dependable in-service-inspection of components to identify
development of cracks or defects and their early repair is even more desirable. Apart
from development of techniques for leak detection, non-destructive examination and in
situ repair we need to be able to assess areas prone to development of defects, assess
growth of defects and the associated safety margin. This requires work to be pursued in
area of stress analysis particularly with regard to residual stresses, fracture mechanics,
fatigue and life assessment etc. Some examples of work done in this area are shown in
figures 10 to 12. Figures 10 and 11 show the stress transients during longitudinal welding of a cylinder and its subsequent annealing. The residual stress distribution after welding and its reduction due to annealing operation can also be seen in these figures. Effect of existence of residual stress on the component life is depicted in figure 12 where it can be seen that with existence of tensile residual stress, the cyclic life gets considerably reduced (compare case 1 with case 2). Further with compressive residual stress.

Figure 10. Stress transients during cylinder welding.

Figure 11. Stress transients during annealing of welded cylinder.
stress the life can be increased (compare case 1 with case 3). This forms the basis of development of various techniques like induction heating for stress improvement, heat sink welding, mechanical pipe lock etc, which are aimed at delaying the component failure.

Even assessment of crack growth and the critical crack length becomes considerably difficult with complications in geometry. Considerable work has been done in this area in view of the assessment need with regard to embrittlement of end-shields in RAPS-I, II and MAPP-I, as well as the feed water nozzles in TAPS.

Another aspect of structural safety pertains to the behaviour of structures to seismic excitations. Figure 13 shows a typical structural model of Dhruva pile block used for seismic analysis. A particular deformed mode shape is shown in figure 14. Although sufficient capability exists in this area, work is required to be pursued to enable nonlinear analysis and water sloshing studies.

9. Secondary shutdown system

The current design of PHWRs starting from Narora reactor incorporates a fast acting secondary shutdown system. The system consists of a number of verticle tubes running through the reactor through which a liquid poison is injected. The number of positions available for this purpose is limited due to severe congestion of various nozzles on top of the calandria. The maximum negative reactivity worth of such a system is thus restricted.

An alternate system which can inject the liquid poison in moderator is therefore considered attractive and needs to be developed. With such a system the tubes running through the core for poison injection can be kept horizontal thereby permitting effective spatial separation between the two shutdown systems (see figure 15). This system is expected to be much simple and not so severely limited in negative reactivity worth.
10. Pressuriser

Pressure control in current PHWRs is maintained by a feed and bleed system. This system can cope with transients to a limited extent and also requires steam side pressure to be regulated depending on the reactor power to minimise the swell/shrinkage on the primary side. The severe fluctuations that are possible with our grids could lead to unacceptable pressure fluctuations causing reactor trip. A need has therefore been felt to incorporate a pressuriser which would damp out the pressure fluctuations and maintain them within the desirable limits. It is expected that incorporation of a pressuriser would lead to more reliable performance of the reactor unit.

We have reviewed various items of engineering development which could lead to improved safety characteristics of the system. The coverage is by no means complete. Further, there are a large number of areas of development which are aimed at improving the performance. Some of the important ones are: (i) improved fuel with
Figure 14. Plot of fourteenth mode of model of Dhruva reactor.

Figure 15. Secondary shutdown system.
graphite lubrication which would minimise pellet clad interaction thereby improving fuel performance particularly under transients; (ii) fuel handling system which is one of the most sophisticated systems and key to successful performance of the reactor plant and (iii) partial boiling in channels which would enable larger power to be extracted through the same reactor unit.

With accumulation of more and more reactor years of operation the need to carry out maintenance as well as some of the planned activities like channel adjustment, channel replacement, steam generator tube plugging, repair on piping system etc would grow. Since these activities require a lot of man-reams it is necessary to develop methods which can enable performance of these activities in minimum time and with minimum of man-rem expenditure.

In conclusion one can say that the existing R & D base can match the needs of the identified power programme. The efforts should be continued to enable further improvements. Some of the thrust areas particularly having a bearing on safety have been identified. These are considered important for realisation of better PHWR system.

References

Atchison R J et al 1983 Nucl. safety 24
Hancox W 1981 Safety research for CANDU reactors, IAEA Tech. Committee Meeting on Thermal Reactor Safety Research, Moscow
Ramanna R 1984 Meghanand Saha Memorial Lecture
Strachan B and Brown D R 1984 Nucl. Eng. Design 81 3
Venkat Raj V, Grover R B, Dolas P K and Mehta S K 1983 Experimental and analytical studies pertaining to small break LOCA in PHWRs IAEA specialist meeting on experimental and modelling aspects of small break Loca, Budapest, Hungary
Accelerator development in India

A S DIVATIA and C AMBASANKARAN*
Variable Energy Cyclotron Centre, Bhabha Atomic Research Centre, Calcutta 700 064, India
* Centre for Advanced Technology, Bhabha Atomic Research Centre, Bombay 400 085, India

Abstract. Accelerator development in India is reviewed with special emphasis on indigenous effort. It started with the 4 MeV cyclotron at the Saha Institute of Nuclear Physics, grew substantially with the installation of the Van de Graaff accelerators at Trombay and really came of age with the 224 cm Variable Energy Cyclotron at Calcutta, which resulted in considerable fall-out of technology. Simultaneously electron linac development has also taken place. Thus the stage is set for developing new types of accelerators such as the electron storage ring synchrotron, and a proton/heavy ion synchrotron.

Keywords. Accelerators; cyclotron; Van de Graaff; Tandem; linac; technology fall-out.

PACS No. 29.15 Br, 29.20 Dt, 29.20 Hm.

1. Introduction

Charged particle accelerator development in the world started in the late 1930s with the primary motivation of seeking fundamental understanding about the structure of the atom. In the last half a century there has been a phenomenal growth all over the world in the field of accelerator technology and a variety of accelerators have been developed for fundamental research as well as applications. It is, therefore, very important to review the development of accelerators and consider the future prospects of accelerator technology in India.

Accelerator development in India started in the 1940s, when a project was undertaken at the Saha Institute of Nuclear Physics, Calcutta to build a 38" cyclotron modelled after the cyclotron at Berkeley (SINP, 1983). Except for the magnet yoke and pole pieces which were imported, all other components were fabricated indigenously. In 1960, it delivered an internal proton beam current of 50–70 μA at 4 MeV and later an external beam current of 0.1 μA.

At the Tata Institute of Fundamental Research, Bombay, some work was done towards building a 1 MeV cyclotron in the 1950s, though it was not made fully operational. At the same time, the 300 keV open air Van de Graaff accelerator was built and operated (George et al 1959) and development work on a 1 MeV electron linear accelerator was carried out.

Simultaneously 400 keV Cockroft-Walton accelerators were built at the Saha Institute and the Bose Institute, Calcutta, a 1 MeV Cockroft-Walton accelerator was bought and installed at the Tata Institute (TIFR 1945–70) and a 150 keV neutron generator was built at the Aligarh Muslim University, Aligarh.

For the first time a large accelerator was available in India when the 5.5 MeV Van de Graaff accelerator purchased from the High Voltage Engineering Corporation (HVEC,
USA) was installed at the Bhabha Atomic Research Centre, Trombay in 1962 (Divatia et al 1962). Though the accelerator was purchased, the switching magnet and the beam transport system were built indigenously. Some time later, a 2 MeV Van de Graaff accelerator, also purchased from the HVEC, was installed at the Indian Institute of Technology (IIT), Kanpur, and a 2 MeV electron Van de Graaff accelerator was built indigenously at the Indian Institute of Science (IISC), Bangalore. Using these accelerators stimulated the scientists and a need for bigger accelerators was felt. HJ Bhabha convened a meeting of scientists on 3 August 1964 at the IISC to consider this question. After the deliberations, he summed up as follows:

"I think the result of this meeting has been extremely instructive and very fruitful. I am convinced that there is a large field of operation in nuclear physics where fruitful work can be done with these machines. The following general situation seems to emerge in my mind. One is that the two machines the Tandem and the AVF accelerator are complimentary. It may be possible for an AVF machine with the help of elaborate equipment to do what a Tandem does. If we decide to go in for a Tandem, it seems clear that we should buy one. The technology of an AVF is entirely different and this is the field, except for the cyclotron at Calcutta, which has been neglected in India. This is the field which I think should now be developed. We are in a position now to enter in a new field of building of accelerators. Therefore, I myself favour the idea of going in for an AVF machine."

2. Variable energy cyclotron at Calcutta

The Atomic Energy Commission decided in 1968 to construct a 224 cm diameter Variable Energy Cyclotron (VEC) at Calcutta indigenously adopting the design of the 88 inch cyclotron at the Lawrence Berkeley Laboratory.

Specifications of the cyclotron are given in table 1. An overall view of the cyclotron and the beam transport lines after construction is seen in figure 1, and a close up of the cyclotron is seen in figure 2. Various stages of progress of the vec have been reported at International Conferences on Cyclotrons and their applications (Divatia et al 1984; Chatterjee et al 1981; Divatia 1979). A brief description of various components of vec and the technologies involved is presented here.

<table>
<thead>
<tr>
<th>Table 1. Design parameters—VEC Calcutta.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Beam</strong></td>
</tr>
<tr>
<td>Maximum energy</td>
</tr>
<tr>
<td>Internal beam current</td>
</tr>
<tr>
<td>External beam current</td>
</tr>
<tr>
<td>Energy resolution</td>
</tr>
<tr>
<td>a) Unanalysed beam</td>
</tr>
<tr>
<td>b) Analysed beam (1 mm slit)</td>
</tr>
<tr>
<td><strong>Magnet</strong></td>
</tr>
<tr>
<td>Pole diameter</td>
</tr>
<tr>
<td>Spiral sections</td>
</tr>
<tr>
<td>Acceleration system</td>
</tr>
<tr>
<td>Frequency range</td>
</tr>
<tr>
<td>Dee voltage</td>
</tr>
<tr>
<td>Extraction system</td>
</tr>
</tbody>
</table>
Figure 1. Overall view of the VEC, Calcutta and the high intensity beam transport lines in the vault.

Figure 2. A close-up of the cyclotron showing the main magnet with the main coil and the first quadrupole magnet near the exit port on the dee chamber.
2.1 Magnet

The most challenging job from the standpoint of indigenous effort was that of constructing the 260 tonne magnet. Since large-sized rolled plates of the requisite quality and size were not available, it was decided to cast the necessary steel blocks, and machine them. The Heavy Engineering Corporation (HEC), Ranchi was the only agency capable of delivering the job, and they agreed to do it. They were highly successful in casting the magnet-quality steel, with carbon content less than 0.1%. Table 2 shows a comparison of the steel composition achieved and that specified. The machining of the magnet, including the 3 spiral-shaped pole tips of the magnet was also carried out successfully to the required accuracy. Figure 3 shows the assembled magnet and figure 4 shows the accuracy achieved in the magnet gap.

<table>
<thead>
<tr>
<th>Specified composition (%) (maximum)</th>
<th>Achieved composition (%) (typical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon</td>
<td>0.12</td>
</tr>
<tr>
<td>Silicon</td>
<td>0.35</td>
</tr>
<tr>
<td>Manganese</td>
<td>0.50</td>
</tr>
<tr>
<td>Phosphorous</td>
<td>0.04</td>
</tr>
<tr>
<td>Sulphur</td>
<td>0.05</td>
</tr>
<tr>
<td>Iron</td>
<td>Balance</td>
</tr>
</tbody>
</table>

Figure 3. The assembled 260 tonne main magnet frame and the main coils during assembly stage.
The main magnet coils, made of etp copper of 29 cm square cross-section with a 1.9 cm hole in the centre for cooling water, providing 5,60,000 ampere turns, were made by the Bharat Heavy Electricals Limited (BHEL), Bhopal to the required accuracy. Similarly the technology of winding and epoxy potting the trim and valley coils was also evolved at the BHEL and fabrication completed to full satisfaction.

The main magnet coils, 17 trim coils and 5 sets of valley coils, all require independent power supplies with stabilized currents. The building of these power supplies was totally an in-house effort, undertaken by the personnel of the vec Centre. For some of the power supplies special water cooled transformers were built. Table 3 shows the specifications of these power supplies and figure 5 shows the inside of a trim coil power supply. The peak power consumption for the main magnet alone is about 1 MW.

Table 3. Power supplies for vec.

<table>
<thead>
<tr>
<th>Unit</th>
<th>No.</th>
<th>Ratings</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main coil</td>
<td>1</td>
<td>150 V, 2800 A</td>
<td>SCR controlled current regulated to 0.01%</td>
</tr>
<tr>
<td>Trim coils</td>
<td>17</td>
<td>6-24 V, 750-2500 A</td>
<td>Series transistor current regulated to 0.01%</td>
</tr>
<tr>
<td>Valley coils</td>
<td>5</td>
<td>15 V, 300 A</td>
<td>Unregulated</td>
</tr>
<tr>
<td>Oscillator</td>
<td>1</td>
<td>20 kV, 20 A</td>
<td>Series tube fast crowbar</td>
</tr>
<tr>
<td>Ion source</td>
<td>1</td>
<td>600 ARC Voltage 500 A filament current</td>
<td>Series tube SCR controlled</td>
</tr>
<tr>
<td>Deflector</td>
<td>2</td>
<td>120 kV, 5 mA</td>
<td>Voltage multiplier</td>
</tr>
<tr>
<td>Switching magnet</td>
<td>2</td>
<td>50 V, 300 A</td>
<td>SCR controlled 0.01%</td>
</tr>
<tr>
<td>Analysing magnet</td>
<td>1</td>
<td>150 V, 500 A</td>
<td>SCR/transistor NMK sensing 0.001%</td>
</tr>
<tr>
<td>Quadrupole magnet</td>
<td>5</td>
<td>30 V, 300 A</td>
<td>Regulation 0.1%</td>
</tr>
</tbody>
</table>
Figure 5. A typical inside view of a regulated power supply.

The total quality of the magnet can best be judged by examining the $B$ vs $I$ curve, shown in figure 6. It can be seen that the curve is in excellent agreement with a similar curve for the Berkeley cyclotron magnet.

2.2 Radiofrequency system

The radiofrequency system of the vec is a 400 kW rf oscillator system in the frequency range 5.5–16.5 MHz, using panel movement for frequency variation. The resonator tank, made from 38 mm thick copper clad steel plates, measuring $2.36 \times 3.13 \times 2.39$ m, was fabricated at the Garden Reach Shipbuilders and Engineers Ltd, Calcutta and the dee, dee stem and panels, made from etp copper, were fabricated at the Central Workshops, barc, Bombay. The rf panels inside the resonator tank are shown in figure 7.
Figure 6. $B$ vs $I$ characteristics of the main magnet; lower portion shows average magnetic field $B$ vs radius.

Figure 7. Radiofrequency movable panels and the dee stem inside the resonator tank.
In the earlier stages, an RF self-excited oscillator system based around the RCA 6949 triode was used. Later on, for reasons of economy and more reliable and stable operation, an RF system of MOPA configuration with the RCA 4648 tetrode has been used. Figure 8 shows a block diagram of the system. The high voltage anode power supply for the RF system was built by the in-house VECC group.

![Block diagram of the radiofrequency system.](image)

**Figure 8.** Block diagram of the radiofrequency system.

### 2.3 Vacuum system

The combined volume of the dee chamber and the resonator tank is 23 m$^3$, and when the vacuum system for evacuating this volume to a pressure in the range of 10$^{-6}$ torr was built, it was the largest one built indigenously. There are two 89 cm oil diffusion pumps, backed by roots and rotary pumps, with proper arrangements of freon cooled chevron baffles and gate valves. The unbaffled pumping speed of each pump is 42,000 litres per second. This pump under initial testing is shown in figure 9. The pumps, baffles and the gate valves were made at the Technical Physics Division, BARC. Subsequently, this technology has been transferred to the Indo-Burma Petroleum Co. and many large vacuum systems have been built by them for the Indian Space Research Organisation (ISRO) and other agencies.

### 2.4 Ion source, probes and deflector

The ion source is a hot cathode pig type and it is mounted on a 3.6 m long shaft which is inserted through a central hole at the bottom of the magnet. There is an air-lock so that the ion source can be pulled out for filament replacement without disturbing the main vacuum. A view of the ion source assembly below the magnet is shown in figure 10. There are three water-cooled copper probes—the target probe, the dee probe and the deflector probe, for monitoring the beam current as a function of the radius, as well as for measuring the radial and vertical characteristics of the beam. All these components are remote controlled from the control room.

The electrostatic deflector has three water-cooled cylindrical electrodes and a tungsten septum. The electrodes can be moved to vary the gap by remote control,
through servo-motors. A 120 kV, 5 mA power supply provides the necessary electrical field in the deflector gap.

The ion source, probes and the deflector were made by the joint efforts of the Central Workshops, BARC and the VECC Workshop.

2.5 Beam transport and data processing systems

The beam transport system consists of three high intensity, low resolution experimental channels which are operational and six high resolution, low intensity experimental channels, which are under installation. There is one switching magnet giving the three high intensity channels, and, after an analysing magnet, another switching magnet giving the six high resolution channels. The steel for the switching and analysing magnets was cast at the HEC, Ranchi and machining has been done in various engineering firms. The coils for the analysing magnet have been fabricated by the BHEL, Bhopal, whereas the coils for the first switching magnet have been made at the VECC
Workshop. Figure 11 shows the first switching magnet under installation. A large number of quadrupole magnets are required for focussing the beam at various stages of the beam transport system. A typical quadrupole magnet is shown in figure 12. The design and fabrication of some quadrupole magnets has been done at VECC and a large number have been fabricated at commercial workshops, under VECC supervision. Many elements of the beam transport system, such as collimating slits, beam viewers, faraday cups, etc have been fabricated entirely at VECC. Figure 13 shows a completed beam transport line in the experimental area.

Facilities for research at VECC include a 915 mm scattering chamber fabricated at VECC, a target facility for making a variety of targets, a detector facility for fabricating various surface barriers and Si(Li) detectors and an electronics facility for producing nuclear instrument modules.
2.6 Low conductivity water system

Since the many magnet coils, power supplies and the rf system, all require extensive cooling, a 1000 gpm, 200 psi, low conductivity (conductivity 1 μmho-cm) water cooling system has been installed. Figure 14 shows the storage tanks and the heat exchangers of the cooling system.
2.7 Control system

Control and monitoring of all the cyclotron systems is done remotely from the control console in the control room, shielded by adequately thick concrete. A view of the control console is shown in figure 15.
2.8 Cyclotron operation and utilization

Regular cyclotron operation has been possible only after the availability of uninterrupted power at VECC from September 1981. The utilization programme has been gradually stepped up. During 1983, the cyclotron was scheduled for operation for 3626 hours; 41% of this time was available for experiments and 17% was available for beam development and tuning; the rest of the time comprised on-line repairs, machine start-up and stopping and grid power failure. \( \text{He}^{++} \) beams in the range of 30 MeV to 75 MeV were utilized. Details regarding beams obtained and utilized are given in table 4. During 1984 the cycle time of cyclotron operation has been further increased.

The utilization of the cyclotron has covered various fields, such as nuclear physics, solid state physics, radiation damage, chemistry and analytical chemistry, radiochemistry, isotope production, biophysics and metallurgy.

As many as 17 research institutions and universities all over the country have already utilized the cyclotron for experiments. They are the Saha Institute of Nuclear Physics, Tata Institute of Fundamental Research, Bhabha Atomic Research Centre, VECC Centre, Reactor Research Centre, Indian Association for the Cultivation of Science, Bangalore University, Mohanlal Sukhadia University, Udaipur, Mysore University, Banaras Hindu University, Aligarh Muslim University, Calcutta University, Kalyani University, Punjab University, Burdwan University, Jadavpur University and IIT, Kanpur.

2.9 Technology fall-out from VECC

There has been considerable fall-out of technology from the project of constructing the Variable Energy Cyclotron indigenously as summarised in table 5. Large industries,
### Table 4. Alpha beam information.

<table>
<thead>
<tr>
<th>Energy (MeV)</th>
<th>Beam current (μA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>35</td>
<td>11</td>
</tr>
<tr>
<td>40</td>
<td>11</td>
</tr>
<tr>
<td>45</td>
<td>10</td>
</tr>
<tr>
<td>50</td>
<td>7.5</td>
</tr>
<tr>
<td>55</td>
<td>5</td>
</tr>
<tr>
<td>60</td>
<td>5</td>
</tr>
<tr>
<td>65</td>
<td>4</td>
</tr>
<tr>
<td>70</td>
<td>2</td>
</tr>
<tr>
<td>75</td>
<td>2</td>
</tr>
<tr>
<td>80</td>
<td>2</td>
</tr>
<tr>
<td>85</td>
<td>1</td>
</tr>
<tr>
<td>100*</td>
<td>2</td>
</tr>
</tbody>
</table>

Energy resolution: 0.5% (FWHM)
Emittance, horizontal: $30\pi$ mm-mrad
vertical: $20\pi$ mm-mrad
Pulse width: $10^\circ$ RF (optimized at 40 MeV)

Status—external in all cases, except where marked with an *.

### Table 5. Fall out of technology from VECC.

- Casting of steel with 0.1% carbon content at HEC, Ranchi for achieving high magnetic fields.
- Epoxy potting of main, trim and valley coils (Cu conductors) of cyclotron at BHEL, Bhopal.
- Precision machining of main magnet at HEC, Ranchi.
- Fabrication of 89 cm dia diffusion pumps with their chevron baffles for the first time in India.
- Fabrication of high voltage, high current, high frequency epoxy cast transformers for use in VECC and other DAE units.
- Fabrication of water-cooled high current transformers.
- Fabrication of high current (3000 A), stabilised (1 in $10^5$) power supplies for cyclotron magnet.
- Fabrication of high power (300 kW), wide band (5–20 MHz) amplifiers.
- Fabrication of precision, high current sensing transformers.
- Fabrication of special magnets such as quadrupole magnets, switching magnets, analysing magnet and steering magnets.
- Development of semiconductor detectors.
- Development of nuclear instrument modules for nuclear physics experiments.

both public and private sector, as well as small industries have been involved in the process, and in-house technological capabilities have been increased while developing new processes.

### 3. Variable energy cyclotron at Chandigarh

The 66 cm Variable Energy Cyclotron at Punjab University, Chandigarh is one of the earlier cyclotrons made, operating around 1953–54 at the University of Rochester,
Rochester, NY, USA (Punjab University 1979). It was transferred first to the Kurukshetra University and then to the Punjab University, where it was installed after modifications and made operational in 1973.

This is a classical single dee cyclotron with arrangements for frequency variation from 10 to 20 MHz and magnetic field upto 14 kiloGauss. Typical external beams available on target are

<table>
<thead>
<tr>
<th>Particle</th>
<th>Current (max)</th>
<th>Energy (MeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protons</td>
<td>2 µA</td>
<td>1-5</td>
</tr>
<tr>
<td>Deuterons</td>
<td>1 µA (max)</td>
<td>1-4</td>
</tr>
<tr>
<td>$^4$He$^{++}$</td>
<td>100 nA (max)</td>
<td>1-2</td>
</tr>
<tr>
<td>$^3$He$^{++}$</td>
<td>100 nA (max)</td>
<td>4-11</td>
</tr>
</tbody>
</table>

4. 2 MV Tandem Van de Graaff accelerator

Operation, maintenance and development of the 5.5 MeV Van de Graaff accelerator at Trombay has generated experience in the technology of electrostatic accelerators. Using this experience, a fully indigenous 2 MV Tandem accelerator has been designed and constructed at Trombay, (Betigiri et al 1982; Singh et al 1982).

There are two support columns, each consisting of 49 sections, each section being separated by 9 unglazed ceramic insulators of 25 mm thickness, bonded by a commercially available two-component adhesive. The pressure vessel contains a $N_2$...
and CO\textsubscript{2} gas mixture at 17.6 kg/cm. A nylon endless belt is used for carrying charge. The accelerating tube consists of index glass rings and stainless steel electrodes. A negative ion source, an einzel lens, a pre-selector magnet, a quadrupole magnet and an analysing magnet, in conjunction with the accelerating tube, comprise the beam handling system. Figure 16 shows the column structure details and the accelerating tube.

The beams obtained are protons: 1 \(\mu\)A; \(^{16}\text{O}^{2+}\): 200 nA; and \(^{16}\text{O}^{3+}\): 40 nA.

5. 4 MeV 400 Rad standing wave linac at Bombay

A 4 MeV 400 Rad standing wave linac has been designed, developed and fabricated at the Tata Institute of Fundamental Research; it has been installed at the Vikram Sarabhai Space Centre (vssc), Trivandrum for x-ray radiography and is in operation since August 1983 (Syunry 1983). A side view of the linac is given in figure 17. The x-ray head is manouverable in all directions so that the x-ray beam can be directed at any part of the heavy fabricated items such as large castings and rocket motors. The operational parameters of the linac are given in table 6.

![Figure 17. A side view of the 4 MeV standing wave linac at TIFR.](image)

6. 8 MeV microtron accelerator, Pune

An 8 MeV microtron has been built and installed at the University of Poona (Asgekar 1980). It gives an internal electron beam current of 10 \(\mu\)A, and an external beam current of 1 \(\mu\)A at an extraction radius of 40 cm. The electron pulses are of 2 \(\mu\)sec duration, and the repetition rate is 50 sec\(^{-1}\).
Table 6. Operational parameters of the 4 MeV standing wave linac at TIFR.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum energy of x-rays</td>
<td>4 MeV</td>
</tr>
<tr>
<td>Dose rate (max)</td>
<td>250 Rad/min at one metre</td>
</tr>
<tr>
<td>Focal spot size</td>
<td>2 mm dia</td>
</tr>
<tr>
<td>Collimator angle</td>
<td>30°</td>
</tr>
<tr>
<td>Field size at one metre</td>
<td>560 mm dia</td>
</tr>
<tr>
<td>Half value thickness</td>
<td>25 mm of steel</td>
</tr>
<tr>
<td>Size of x-ray head</td>
<td>$760 \times 760 \times 1700$ mm</td>
</tr>
<tr>
<td>Weight of x-ray head</td>
<td>1000 kg</td>
</tr>
<tr>
<td>Beam centering</td>
<td>By laser beam direction finder</td>
</tr>
<tr>
<td>Movement of x-ray head</td>
<td>Telescopic vertical up to 5 m;</td>
</tr>
<tr>
<td></td>
<td>$\pm 175^\circ$ about vertical axis;</td>
</tr>
<tr>
<td></td>
<td>$+45^\circ$ to $-95^\circ$ about horizontal axis;</td>
</tr>
<tr>
<td></td>
<td>Long and cross travel by EOT crane</td>
</tr>
</tbody>
</table>

7. Future accelerators

A 2 MV Tandem accelerator, exactly similar to the 2 MV Tandem accelerator at Trombay, is being fabricated at the Reactor Research Centre, Kalpakkam, and it is expected to be commissioned in 1985.

A 14 MV Pelletron purchased from the National Electrostatic Company (NEC), USA, is being installed at the Tata Institute of Fundamental Research, Bombay by BARC and TIFR. It will accelerate protons to 28 MeV, alphas to 42 MeV and heavy ions such as $^{28}$Si, $^{32}$S, $^{40}$A and $^{64}$Ni. This accelerator is expected to be operational in 1985. A superconducting linac booster is planned as a post-accelerator.

A major programme of accelerator development is envisaged at the Centre for Advanced Technology, BARC, Indore. Proposals are being studied for an electron storage ring synchrotron, a proton/heavy ion synchrotron (Jain and Divatia 1977) and a medical cyclotron.

8. Conclusion

A variety of accelerators have been designed, developed and built in India. Wherever accelerators have been purchased, associated development work has been necessary. Construction of the Variable Energy Cyclotron at Calcutta has given a major boost to the accelerator technology and established a strong liaison between Research Institutes and the Industry.

Besides the accelerators discussed here, there are a number of accelerators, with energy less than 1 MeV, in Research Institutes and universities, which are used as Neutron Sources or for ion implantation. Quite a few medical institutes and hospitals have procured electron accelerators for radiation therapy. The need for accelerators and development of accelerator technology is growing.

Accelerator development in India has been pioneered by Meghnad Saha, given a major thrust with a vision by Homi J Bhabha, encouraged by Vikram A Sarabhai and brought to a high technology level with a promise for the future by Raja Ramanna.
Acknowledgements

We are grateful to the Department of Science and Technology, Government of India, for information about various accelerators in India. We wish to thank Shri R K Bhandari and Shri N K Mukhopadhyay for assistance.

References

Chatterjee S et al 1981 IXth Int. Conf. on Cyclotrons and their Applications, (Caen France: Proceedings) 75
Divatia A S et al 1962 AEET/NP/5
Jain A and Divatia A S 1977 8 110
Punjab University 1979 Report on Chandigarh Variable Energy Cyclotron
Singh P, David T P and Betigiri M G 1982 Pramana 19 337
SINP 1983 30 years of Saha Institute of Nuclear Physics, 1951–1981. Rep., p. 27
Syunry T 1983 Nuclear India 21 (12)
Phenomenology of the Pokaran PNE experiment

R CHIDAMBARAM, S K SIKKA and SATISH C GUPTA
Physics Group, Bhabha Atomic Research Centre, Trombay, Bombay 400 085, India

Abstract. The phenomenology of the Pokaran PNE experiment (yield - 12 kiloton of TNT) conducted in a shale-sandstone rock, 107 meters underground, is described with the aid of computations using a one-dimensional spherical symmetric rock mechanics computer code developed by the authors. The calculated values of cavity radius, spall velocity and extent of rock fracturing are in good agreement with the observed values. The principal mechanism for crater formation at Pokaran was spall and the relatively smaller crater dimensions and non-venting of radioactivity gases were due to lower kinetic energy transferred to the shale-sandstone rock.

Keywords. Pokaran; peaceful nuclear explosions; crater; non-venting.

PACS No. 91-90; 91-60.

1. Introduction

India's first and only peaceful nuclear explosion experiment was carried out in Rajasthan desert on 18 May 1974 at a place near Pokaran. The object of the experiment was to study the explosion phenomenology, fracturing effects in rocks, ground motion, containment of radioactivity and postshot access of the region around the point of detonation (Chidambaram and Ramanna 1975). Such studies are important in the context of possible applications of peaceful nuclear explosions (PNEs).

A plutonium device, of yield 12 kt equivalent of TNT,* was emplaced in a shale medium, at a depth of 107 m, in a chamber at the end of an L-shaped hole. Upon detonation, the ground surface above the emplacement point rose with a velocity of about 25-30 m/sec to form a dome 170 m in diameter and 34 m in height. The mound remained intact during its growth and fall and no venting of radioactive gases was observed. The resultant apparent crater, measured with respect to the preshot ground surface, has an average radius of 47 m and depth of 10 m.

The phenomenology of the event is interesting as it is the only recorded case of an explosion which produced a crater, although a very shallow one, and yet was completely contained from the radioactivity point of view. Also, a comparison of the Pokaran crater dimensions with other cratering experiments in hard rock (figures 1 and 2) shows that the Pokaran crater was smaller. At its scaled depth of burial (SDOB) 51.5 m/kt$^{1/3.4}$, the scaled radius and depth for a hard rock cratering experiment are predicted to be (Nordyke 1970) SR$_d$ = 27 m/kt$^{1/3.4}$ and SD$_d$ = 14 m/kt$^{1/3.4}$, compared to the values of 23 m/kt$^{1/3.4}$ and 5 m/kt$^{1/3.4}$ respectively observed for the Pokaran experiment. The

* 1 kt = 4.186 x 10$^9$ ergs, the explosive energy equivalent of 1000 tonnes of TNT.
peak vertical velocity of 25 to 30 m/sec for the latter also is lower than the value \( \sim 40 \) m/sec, as read from figure 6 of Toman (1970). In this paper, we have tried to explain these observations based on computer calculations using 'OCENER', a one-dimensional spherical rock mechanics code developed in our laboratory (Gupta et al 1979a; Gupta 1979).
2. One-dimensional rock mechanics code

The computer code developed by us has many similarities with the American code 'soc' (Cherry 1967; Cherry and Petersen 1970; Schatz 1973) and the French code 'S' (Michaud and Maury 1970). Some of the physical-mechanical processes associated with the propagation of the stress field set up in a geological medium by a sudden release of the explosive energy of a nuclear device (Teller et al 1968) are vaporisation, melting, crushing, fracture and motion of the rock. Also the stress wave gets reflected at the free ground surface and this transfers additional kinetic energy to the rock medium. The final crater dimensions depend upon the total kinetic energy transferred to the region above the explosion-produced cavity. Our code models these events in one dimension, in spherical symmetry, by finite difference solutions of the laws of mass, momentum and energy conservation, and employs the physical and mechanical properties of the rock material under different stress conditions.

In a given calculation, the medium surrounding the explosion point is divided into spherical meshes and the conservation laws are expressed in a Lagrangian co-ordinate system, so that the conservation of mass is automatically achieved. From the initial stress field at time $t$, the momentum equation (equation of motion) is solved to ascertain the acceleration of each zone boundary. Accelerations, when allowed to act over a small time increment $\Delta t$, produce new velocities and subsequently new displacements of the zone boundaries. These displacements produce strains which, depending upon the equation of state of the rock, produce the new stress field. The cycle is repeated after each time increment $\Delta t$. The steep shock fronts are handled in the program by the artificial viscosity method due to von Neumann and Richtmayer (1950). Some of the outputs of this program are the following:

(i) the amount of shock-vapourised rock,

(ii) the amount of shock-melted rock,

(iii) the radius of the cavity and the pressure inside it at any given time,

(iv) profiles of shock position versus time, peak stress (both hydrostatic and deviatoric) versus radius, particle velocity versus radius, etc.,

(v) spall velocities of the mesh points as the reflected tensile wave from the surface travels towards the growing cavity,

(vi) kinetic energy transferred to the medium, and

(vii) the extent and type of rock failure (ductile or fracture).

More details of the code are given elsewhere (Gupta et al 1979a).

3. Medium properties*

The geology around the emplacement point has been briefly described by Chidambaram and Ramanna (1975). The shot point was located in a sedimentary

* The measurements of the various medium properties reported in this section were carried out in several divisions of BARC and elsewhere.
formation and the area surrounding the surface ground zero was a flat terrain covered with sand and sandy loam type of soil. The section of the geological structure adopted for the calculations is shown in figure 3. Some of the measured physical and mechanical properties are also shown. The chemical composition of the rocks suggests that the major constituent of these rocks is SiO$_2$ and the rock can be classified as 'sedimentary SiO$_2$ rock type'.

### 3.1 P-V curves

The experimental shock Hugoniot data for Pokaran shale* are plotted in figure 4 along with data for several shales and sandstones† found in USA. It can be seen that an 'average' curve for a shale medium can be drawn independent of initial density and hence of porosity. This implies that the compressibility at higher pressures does not vary much with minor variations of the rock constituents and is a function of grain density only. This appears to be reasonable since with increasing pressure, the pore collapse and porosity disappears irreversibly at about 40 kbar (Stephens et al 1969). Such averaging of the Hugoniot data for calculations for parameters studies using computer codes has also been done by Hearst (1971) for alluvium and by Butkovich (1973) for granite, basalt and tuff rocks. It is interesting to note that the Hugoniot curve calculated for Pokaran shale ($\rho = 2.45 \text{ g/cm}^3$) using the Butkovich's (1973) prescription, matches well with the 'average' curve. The P-V curve for Pokaran shale, therefore,\

---

* The sample for these measurements had density of 2.22 g/cm$^3$.
† These data for other rocks are taken from figure 2 of Terhune (1973).
The Pokaran PNE experiment

Figure 4. Shock Hugoniot for Pokaran shale compared with various sandstones and shales. Full curve is the one calculated for \( p_0 = 2.45 \text{ g/cm}^3 \) using **pmugen**.

was calculated, including the lower pressure range (below 40 kbar), employing a Trombay program **pmugen** based on Butkovich's prescription (Gupta 1979). For Pokaran sandstone, since the shock wave decays to less than 5 kbar when it reaches shale-sandstone interface, the P-V curve is required only in the low pressure regime. This also is calculated using the program **pmugen**.

3.2 Strength curves

Since the facilities to carry out triaxial tests on rock samples were not available, the measured values of unconfined compressive strength of the rock have been used to estimate the strength curves employing the expression given by Gupta and Sikka (1978).

\[
\frac{C}{C_0} = 1 + k\left(\frac{\sigma_3}{C_0}\right)^n,
\]

where \( C \) is the confined compressive strength at the confining pressure of \( \sigma_3 \) and \( C_0 \) is the unconfined compressive strength (\( \sigma_3 = 0 \)). \( k \) and \( n \) are constants for a given rock type. The values \( k = 2.1 \) and \( n = 0.64 \) for shales and \( k = 3.2 \) and \( n = 0.67 \) for sandstones were employed. The brittle ductile points are placed at

\[
J_m = 0.8P_m,
\]

following Mogi (1966) for silicate rocks, where \( J_m \) is the maximum shear stress and \( P_m \) is the mean pressure.
Table 1. Average physical properties of the shale and sandstone rocks used for calculations.

<table>
<thead>
<tr>
<th>Property</th>
<th>Shale</th>
<th>Sandstone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density ($\rho_0$)</td>
<td>2.45 g/cm$^3$</td>
<td>2.10 g/cm$^3$</td>
</tr>
<tr>
<td>Porosity ($\phi^*$)</td>
<td>12%</td>
<td>25%</td>
</tr>
<tr>
<td>Water content</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>Unconfined compressive strength (bar)</td>
<td>256</td>
<td>60</td>
</tr>
<tr>
<td>Tensile strength (bar)</td>
<td>30</td>
<td>19</td>
</tr>
<tr>
<td>Ultrasonic pulse velocity (km/sec)</td>
<td>2.2-95</td>
<td>1.4-2.7</td>
</tr>
<tr>
<td>Young's modulus (kg/cm$^2$)</td>
<td>$0.8 \times 10^5$</td>
<td>$0.5 \times 10^5$</td>
</tr>
<tr>
<td>Poisson's ratio</td>
<td>0.10</td>
<td>0.28</td>
</tr>
</tbody>
</table>

* $\phi = (\rho_G - \rho_0)/\rho_G$, where $\rho_G$ is the grain density and $\rho_0$ is the bulk density of the rock.

The average physical and mechanical properties of shale and sandstone rocks, used in the calculation, are given in table 1.

4. Calculations and results

The total energy of the device, 12 kt ($= 5.04 \times 10^{20}$ ergs), was released in less than one microsecond into the device material. The resulting temperature and pressure were estimated, following Teller et al (1968), to be ~ 4 million degrees K and ~ 70 Mbar respectively. At this temperature, the device materials would have got completely vaporized. This gas then expanded rapidly and adiabatically to fill the emplacement chamber.

4.1 Shock vaporisation and shock melting

To determine the shock-vaporised and shock-melted regions of rock, a fine mesh calculation was done. The energy of the device ($W = 12$ kt) was distributed uniformly as internal energy of sphere of 'iron gas', whose volume is equal to that of the emplacement chamber. The vaporisation phase was found to be completed at about 150 $\mu$s and the total mass of the rock vaporised was 640 tons which extends up to a radius of 4.1 m from the shot point. About 2000 tons of the rock, extending up to a radial distance of 6.2 m, was shock melted in about 600 $\mu$s. The amounts of rock shock vaporised and shock-melted are 53 tons/kt and 160 tons/kt respectively, which are lower than the values of 70 tons/kt and 300 tons/kt respectively calculated for USA hard rock events (Butkovich 1967).

4.2 Shock propagation and cavity growth

These calculations were performed using the so-called 'bubble model' (Butkovich 1967). The total energy of 12 kt was redistributed uniformly in the spherical mass of
vaporised rock of radius 41 m. At this energy density, the spherical cavity had uniform pressure of about 1.6 megabars. The equation of state used for the vaporised rock was the one for a \( \text{SiO}_2 + 1\% \text{H}_2\text{O} \) gas mixture taken from Butkovich (1967).

The position of the first shock, the rarefaction wave and the cavity growth are plotted in figure 5. The shock position was taken to be the point where the magnitude of the artificial viscosity term was greatest. At about 14 msec the shock wave reached the shale-sandstone interface. At this interface a shock wave was induced in the sandstone layer while a backward moving rarefaction wave travelled in the shale medium*. The shock wave in sandstone split at about 62 m from shot point into two waves: a main shock wave and a preceding elastic precursor which moved with the sound velocity in the medium. It can be seen that the main shock reached the free ground surface at time \( t_s \approx 41 \) msec and got reflected from there as a rarefaction wave and met the top of the growing cavity at \( t_g = 72 \) msec.

The radius of the cavity at time \( t_g \), in the vertical direction was 24.5 m compared to 23.5 m in a horizontal direction (all-shale calculation). This difference is caused by the rarefaction wave produced by the reflection of the first shock at the shale-sandstone interface. As upto time \( t_g \) (Rodinov et al 1971) the cavity radius is able to attain 80–90%.

* From this time onwards, the one dimensional symmetry of the problem is lost and the calculations using 'ocener' are strictly valid for the vertical direction connecting the shot point and the surface ground zero.
of its final value, the final cavity radius in horizontal direction is estimated to be 28–29.5 m. This is in comparison with post-shot measured value of 30 m.

4.3 Particle velocity profiles

The particle velocity versus time at distances 27.4 m, 55 m, 80.5 m and 107 m from the shot points are plotted in figure 6. The first outward going shock wave (indicated by the sharp increase in the particle velocity) appears at these distances around 6.4, 16, 29 and 41 msec respectively. Surface ground zero begins to move up with initial velocity of 30 m/sec. This is in good agreement with the observed velocity (25–30 m/sec).

The rarefaction wave starts from the surface ground zero and moves towards the cavity increasing the particle velocities (indicated by points A in figure 6). The plot shows that the shale-sandstone interface spalled at around 58 msec. A recompaction wave, which resulted in further increase in particle velocities (denoted by points B), seems to be arrested at the shale-sandstone boundary. This would have reached the surface at much later times and its effect on increasing the particle velocities (and hence

![Figure 6. Computed particle velocity versus time profiles at various distances for the Pokaran PNE experiment. A — spall phase, B — gas acceleration phase.](image)
on the final crater dimensions) would seem minimal. This calculation seems to suggest that the principal mechanism responsible for crater formation at Pokaran was spall.

4.4 Extent of rock failures (ductile flow and brittle fracture)

The calculation shows that all the rock above the detonation point (i.e. in vertical direction) had failed. In horizontal direction the extent of fracturing was determined by performing a calculation in homogeneous shale medium in which the free surface had been removed far away from detonation point. The maximum extent of fracture radius, as indicated by the crack number equal to 2, is about 114 m compared to the observed value of 80–100 m.

4.5 Reasons for smaller crater in Pokaran experiment

Having described the early time phenomenology of the Pokaran experiment, we shall in this section examine the causes for obtaining a smaller crater compared to the cratering events in hard rocks in USA. For this we will use the kinetic energy criteria as given in Appendix A.

4.5a Comparative study of cratering in Pokaran medium and Buckboard-basalt*—First, to get a preliminary idea about the cratering efficiency of the Pokaran rocks, we compare in table 2 their physical properties with those of Buckboard basalt in which USA has carried out two cratering-type nuclear explosions: Danny Boy (yield = 0.42 kt) and Sulky (0.087 kt). The USA scaled apparent crater radius and scaled apparent crater depth versus scaled depth of burst curves are essentially based on these experiments. The influence of the variation of rock properties on the cratering efficiency of a rock has been discussed by Terhune et al (1971) and Gupta et al (1979b). Based on that, we have included a comment about the relative cratering efficiency of Pokaran rocks with respect to basalt. We observe that except for the water content the other properties will lead to lowering the cratering efficiency of Pokaran rocks.

<table>
<thead>
<tr>
<th>Property</th>
<th>Shale</th>
<th>Sandstone</th>
<th>Basalt</th>
<th>Resultant cratering efficiency of Pokaran rocks relative to basalt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density (g/cm³)</td>
<td>2.45</td>
<td>2.10</td>
<td>2.62</td>
<td>Lower</td>
</tr>
<tr>
<td>Porosity (%)</td>
<td>12</td>
<td>25</td>
<td>3</td>
<td>Lower</td>
</tr>
<tr>
<td>Bulk modulus (kbar)</td>
<td>62</td>
<td>52</td>
<td>90</td>
<td>Lower</td>
</tr>
<tr>
<td>Water content</td>
<td>0.04</td>
<td>0.02</td>
<td>0.0</td>
<td>Higher</td>
</tr>
<tr>
<td>Maximum shear strength (bar)</td>
<td>358</td>
<td>170</td>
<td>300</td>
<td>Lower</td>
</tr>
</tbody>
</table>

* The physical properties for Buckboard-basalt have been taken from Cherry (1967).
For a quantitative comparison, we performed a calculation for a hypothetical event in which a device of 12 kt yield is detonated at 107 m in Buckboard basalt medium. Figure 7 shows the particle velocity versus range profile at time $t_g$ for Pokaran event and the event in Buckboard basalt. In general, the particle velocities in Pokaran medium are lower than those in Buckboard basalt. The spall velocity, as expected, is also lower (30 m/sec compared to 36 m/sec). Accordingly, the cratering efficiency of Pokaran rock should be lower. This is also indicated by the values of the late-time kinetic energies along the vertical direction. These are 5.7% and 9% respectively in the two cases. Following Burton et al (1975), this gives us an effective equivalent yield of $5.7/9 \times 12 = 7.6$ kt for a detonation at 107 m in Buckboard basalt, which would give similar crater dimensions as the Pokaran experiment. Actually, this may be slightly a lower estimate of the effective yield as for other than vertical directions, the shock wave will traverse more through the shale medium which, as an all shale calculation shows, will have more kinetic energy imparted to it. For all shale calculations, the kinetic energy at time $t_g$ is 6.7% with corresponding effective yield of 8.9 kt. Further, there will be the effect due to interaction of shock waves at the interface of the shale-sandstone medium, which will redirect the particle velocity vectors from their initial radial directions. This is not accounted for in one-dimensional calculations. A rough estimate of this puts the effective SDOB of Pokaran experiment to be $57 \text{ m/kt}^{1/3 - 4}$.

Considering all the above factors, we may say that the Pokaran experiment was equivalent to that of an ~ 8 kt detonation at the same depth (effective scaled depth of burst = 58 m/kt$^{1/3 - 4}$) in basalt. It is known that at this range of scaled depth of burial one gets a retarc or a marginal crater depending upon the bulking properties of the mound fall-back material. Since the mound in Pokaran experiment did not break and
also the post-shot analysis showed that no significant bulking of the rock had taken place, the formation of a marginal crater is not surprising.

In conclusion, the reason for obtaining a smaller crater at Pokaran is due to the fact that a smaller amount of the device energy was converted by the medium into kinetic energy. This is principally due to the geological properties of the Pokaran medium.

4.6 Non-venting of Pokaran experiment

According to the published radioactive vent-fraction $f_c$ versus scaled depth of burial curve (Lapage 1974; Siddons 1974), for the Pokaran experiment at $sdob$ of 51.5 m/(kt)$^{1/3}$.4, $f_c$ is predicted to be very small but to have a finite value of 0.003, instead of the observed value of zero. We feel that this small discrepancy is because such scaling laws ignore the fact that the vent fraction depends not only on $sdob$ but also on the type of the rock medium in which the experiment is carried out. Now, as discussed above, the physico-mechanical properties of a rock reflect themselves well in the amount of energy of the explosion imparted to the rock as kinetic energy and the effective yield of an explosion relative to a standard rock medium can be defined as

$$W = x W,$$  

where $x$ is the ratio of the kinetic energies given to rock medium and a standard rock medium by explosions of yield $W$ carried out at the same depth of burst $d$. This allows us to simply modify the scaling law for $f_c$ from

$$f_c = f(d/(W)^{1/3}),$$  

(4)

to

$$f_c = f(d/(x W)^{1/3}),$$  

(5)

we take

$$x = v_x/v,$$  

(6)

where $v_x$, and $v$ are the peak velocities of surface ground zero for the two explosions in a given medium and the standard medium. For quantitative estimates, we have again chosen buckboard Mesa basalt rock in USA as standard. The variation of $v$ with $d$ for this rock type (see Harlan 1967) is

$$v = 19700 (d/W^{1/3})^{-1.6}.$$  

(7)

Here $W$ is in kilotons and $d$ is in meters. For Pokaran event, $x$ turns out to be 0.5 compared to 1.0 for basalt. Using this, $f_c$ predicted for Pokaran event is $< 10^{-6} \approx$ zero in agreement with experiment.

To summarize, the non-venting of the Pokaran experiment seems to have been due to the smaller cratering efficiency of the shale-sandstone medium. Because only a small kinetic energy was imparted to the mound, it did not rise very high relative to the depth of burial. Consequently, the tensile stresses developed in the mound were not enough to disintegrate it and cause dynamic venting. Also, because the mound did not rise much, the minimum thickness of the overburden layer above the cavity was much larger ($\sim 24$ m compared to 2 m for the Danny Boy event which vented with $f_c = 0.042$) so that there was no connection of fissures. This prevented the radioactivity from leaking into the atmosphere.
Appendix A. On the correlation between the crater dimensions and the mound kinetic energy for peaceful underground explosions

The terminal effect of an underground peaceful nuclear explosion on a geological medium is to produce a crater if the depth of the burial of the explosive device is less than about $60 \text{ m/kt}^{\frac{1}{3.4}}$ (Teller et al. 1968). The dimensions of such a crater, as indicated by many computer code calculations (Terhune et al. 1970; Burton et al. 1975), depend on the motion of the mound which in turn depends on the fraction of total energy of the explosive converted into the kinetic energy of the mound. Any change in the explosion conditions such as the type of the explosive, chemical or nuclear, variation in the rock properties, such as moisture content, compressibility, strength etc or geological variations such as layering, will get reflected in the amount of energy transferred to the mound and affect the crater dimensions. Although the dependence on the kinetic energy was recognised quite some time ago, no quantitative relation appears to have been deduced between it and crater dimensions. Here we attempt to show that such a relation exists.

To drive the relationship, we have carried out a similarity analysis of the linear crater dimensions versus kinetic energy of the mound following the procedure given by Baker et al. (1973). A linear crater dimension $l$ ($l = \text{apparent crater radius} R_a$ or apparent crater depth $D_a$) produced by an explosion, beneath the surface of earth is at least a function of the explosive yield ($W$) and depth of burst ($d$). We assume that the effect of the additional variables like the type of explosive source (chemical or nuclear) and geological medium (compressibility, porosity, strength, water content, etc) is reflected in the kinetic energy ($E$) developed in the mound. This, we represent by a function $l \sim (d, W, E)$.

It is then easy to show from dimensional analysis that

$$l/d = f(E/W).$$

This relation means that if for two craters,

$$E_1/W_1 = E_2/W_2.$$  

Then

$$l_1/d_1 = l_2/d_2,$$

i.e. the two craters will be similar.

Though this analysis gives the functional dependence of the linear crater dimensions on the kinetic energy, the form of the function is yet undetermined. In figures A1 and A2, we have plotted $R_a/d$ versus $E/W$ and $D_a/d$ versus $E/W$ respectively for chemical explosions in Beerpaw shale. The data, which have been taken from Burton et al. (1975), include the values of the quantities $R_a$ and $D_a$ and kinetic energy $E$, which has been calculated using computer code TENSOR (Maenchen and Sack 1964). Although there are only four points in the plots, the correlation between $R_a/d$ and $E/W$ and between $D_a/d$ and $E/W$ are clearly brought out by this data. A linear relationship is implied. It may be noted that the kinetic energy of the mound is not an observable quantity and also is not easily computable as it requires two-dimensional codes like TENSOR (Maenchen and

* This value depends upon the geological medium.
Sack 1964). Hence for practical utilization of a crater dimension versus $E/W$ correlation, we have to have a suitable measure for this kinetic energy. Briefly, the various equivalent measures adopted for this mound kinetic energy, as summarized by Gupta et al (1979b) are:

1. velocity profile between the cavity and the free ground surface in the vertical direction at the time the rarefaction arrives at the cavity surface;
2. free-field kinetic energy versus time profile from a calculation in which the free ground surface is not included in the calculation; and
3. peak pressure versus radius profile.
References


Burton D E, Snell C M and Bryan J B 1975 *Nucl. Technol.* 26 65

Butkovich T R 1967 Lawrence Livermore Laboratory, Livermore Rept. UCRL-14729

Butkovich T R 1973 Lawrence Livermore Laboratory, Livermore Rept. UCRL-51441


Gupta S C and Sikka S K 1978 BARC Rept. BARC-968

Gupta S C 1979 Ph.D Thesis, Bombay University

Gupta S C, Sikka S K and Chidambaram R 1979a BARC Rept. BARC-1023


Harlan R W 1967 Project Pre-GONDOLA I Final Report, PNE-1107

Hearst J R 1971 Lawrence Livermore Laboratory, Livermore Rept. UCID-15783

Lapage R 1974 AWRE Rept. 011/74


Mogi K 1966 *Bull. Earthquake Res. India* 44 215


Redinov V N, Aduskin V V, Kostiuchenko V N, Nikolaevsku V N, Romashov A N and Tavetkov V N 1971 Lawrence Livermore Laboratory, Livermore Report, UCRL-Trans-10671

Schatz J F 1973 Lawrence Livermore Laboratory, Livermore Rept. UCRL-51352


Stephens D R, Lilley E M and Louis H 1969 Lawrence Livermore Laboratory, Livermore Rept., UCRL-71238


Terhune R W 1973 Lawrence Livermore Laboratory, Livermore Rept. UCRL-91377


Perspectives in neutron physics research

P K IYENGAR and K SUBBA RAO
Bhabha Atomic Research Centre, Trombay, Bombay 400085, India

Abstract. Discovery of the neutron in 1932 by Chadwick ushered in a new era of scientific research and technology. The neutron is endowed with unique properties in its mass, life time, spin and magnetic moment etc and every important property has been used in the study of condensed matter, biological molecules, nuclear forces, stellar objects and other fields. Neutron has a wide range of applications in power production, breeding of fissile fuel, radiography, medicine and others.

Keywords. Neutron properties; spallation; plasma focus; breeding; neutron radiography.

PACS No. 28-40; 29-25; 52-75

1. Introduction

It is seldom realised how important the discovery of neutron has been in our understanding of nature. Neutron by virtue of being a major constituent of the nucleus has been responsible for the subject of nuclear physics. Neutrons have not only been the key for nuclear transformations, production of isotopes, production of nuclear power and the like but are also fundamentally responsible for the characteristics displayed by the sun and the stars as we see them today.

In India Dr Raja Ramanna's name is synonymous with neutron physics. Having started his research career in King's college London using a small Ra-Be neutron source for fission physics studies, he has emerged as a pioneer in nurturing the field of neutron physics in our country. He set the pace in early fifties for the Indian scientists to do frontline research in the then emerging fields like solid state physics, nuclear physics, neutron physics and reactor physics using neutron as a research tool. He headed the groups that were responsible for providing high intensity neutron sources from reactors. The reactors APSARA, CIRUS, PURNIMA and latest DHRUVA have all been built under his guidance. India's first peaceful nuclear explosion was also his brainchild.

Neutron plays a very pivotal role as a research tool through some of its unique properties like charge neutrality, spin, magnetic moment etc. Therefore the first part of this review article is devoted to highlight these various properties. The second part of this review details the various techniques employed for the production of neutrons. It is emphasised that it is not merely the intensity alone but rather combination of number and energy that imparts special characteristics. Two applications have been cited at the end in some detail.

This article is dedicated to Dr Ramanna on the occasion of his 60th birthday but for whose vision and missionary zeal neutron physics would not have taken roots and grown to attain its present stature in India.
2. Discovery of the neutron

Lord Rutherford, the discoverer of the nuclear atom, made an amazing prophecy (Rutherford 1920) by stating

"Under some conditions it may be possible for an electron to combine much more closely with hydrogen nucleus (than in the neutral hydrogen atom) forming a kind of neutral doublet. Such an atom would have very novel properties. Its external field would be practically zero, except very close to the nucleus, and in consequence it should be able to move freely through matter. Its presence would probably be difficult to detect by the spectroscope and may be impossible to contain in a sealed vessel. On the other hand it should enter readily the structure of atom and may either unite with the nucleus or be disintegrated by its intense field, resulting possibly in the escape of a charged hydrogen atom or an electron or both."

The discovery of the neutron by Chadwick (1932) twelve years after the prophecy was the culmination of various investigations in several laboratories in Europe. It all started in 1930 when Bothe and Becker (1930) tried to detect nuclear disintegration with a needle counter. They bombarded various light elements like Li, Be and B with alpha particles from a Polonium source and measured the energies of gamma rays emitted from the samples. The gammas from Li were ascribed to nuclear excitation by the inelastic scattering of the alpha particles and that from boron was attributed to the fact that the residual carbon nucleus is in an excited state. Beryllium in particular gave gammas of very high energies of \(~ 5\) MeV, and this was explained by Bothe and Becker saying that the nucleus was excited by the capture of alpha particle.

Two years later, Curie and Joliot used a much more powerful Pu-Be source to investigate further the Be + \(\alpha\) reaction. They observed that the radiation from Be ejected protons with kinetic energies around 5 MeV, when the foils containing hydrogenous material were inserted between beryllium and the ionization chamber. This lead to a difficult conclusion that in order to produce such energetic compton recoils, the gammas must have energies of the order of 50 MeV.

Shortly thereafter, in 1932, Chadwick who was a student of Lord Rutherford, measured energies and momenta of recoils from different materials using a pulse ionization chamber (figure 1) and explained all observations in terms of collisions, consistent with the laws of conservation, with a penetrating neutral massive particle slightly heavier than the proton. He christened the new particle the "neutron" \((n)\) and its production by the bombardment of alpha particles on beryllium was described by the equation

\[
\text{Be}^9 + \text{He}^4 \rightarrow \text{C}^{12} + n.
\]

![Figure 1. Experimental set-up used by Chadwick that led to the discovery of the neutron.](image-url)
The discovery of neutrons resolved many puzzling questions vis-a-vis nuclear structure. Heisenberg immediately propounded that the nuclei contain only protons and neutrons. The discovery of the neutron gave birth to a new branch of physics called neutron physics.

3. Properties of neutrons

3.1 Mass of the neutron

Chadwick (1932) himself concluded from the study of recoil protons liberated by the neutrons, that the mass of the neutron is approximately equal to that of the proton, probably between 1.005 and 1.008. A more accurate estimation of the neutron mass (Chadwick and Goldhaber 1934) was made from the binding energy of the deuteron. The mass of the deuteron is equal to the sum of neutron mass \( m_n \) and proton mass \( m_p \) less the mass defect corresponding to the binding energy \( E_d \). Thus

\[
m_n = m_d - m_p + E_d/c^2.
\]

From the known values of \( m_d (= 2.013554 \text{ amu}) \), \( m_p (= 1.007226 \text{ amu}) \) and \( E_d (= 2.225 \text{ MeV}) \), the neutron mass turns out to be 1.008665 amu. Measurement of the photon energy emitted when slow neutrons are captured by hydrogen lead to further improvement in the neutron mass measurement (Knowles 1962). The accepted value (1982) for the rest mass of neutron is

\[
m_n = 1.008665012 \text{ amu}.
\]

An interesting experiment to measure the neutron mass directly with free neutrons was proposed by Stedman (1968) and several others. A neutron moving with a velocity \( v \) is associated with it a wavelength given by

\[
\lambda = h/(m_n v).
\]

It is clear from this relation that \( h/m_n \) can be determined by measuring the velocity and wavelength of the neutrons. Time of flight technique was proposed to measure the velocity and the wavelength can be determined by Bragg reflection of the same neutrons in a single crystal.

3.2 Charge of the neutron

There is an abundance of evidence that the electric charge on the neutron if any \( (q_n) \) is approximately zero, as noted by Chadwick himself. Fermi and Marshall (1947) obtained an upper limit for the charge as \( 10^{-18} \text{ e} \) from neutron scattering experiments on xenon. Shapiro and Estulin (1956) tried to deflect neutrons using strong electric fields but could not find any effect. Shull et al (1967) obtained an accurate value for the ratio of the neutron charge to that of the proton, using double crystal spectrometer, as

\[
( -1.9 \pm 3.7 ) \times 10^{-18}.
\]

In a recent experiment at the Institut Laue-Langevin at Grenoble Gohler et al (1982) used a focussed beam of 220 m/sec neutrons to pass through an electric field of
5.9 kV/mm for a distance of 10 m and estimated the charge of the neutron as
\[ q_n = -(1.5 \pm 2.2) \times 10^{-20} e, \]
where \(e\) is proton charge.

### 3.3 Spin of the neutron

The deuteron has spin 1 and no orbital angular momentum. Since the proton has spin \(\frac{1}{2}\), it follows that the neutron must have spin \(\frac{1}{2}\) or \(\frac{3}{2}\). Schwinger (1937) showed that experiments on the scattering of neutrons by ortho and parahydrogen would permit determination of the neutron spin. Such experiments carried out by Halpern et al (1937) conclusively demonstrated that the neutron spin is just \(\frac{1}{2}\).

### 3.4 The magnetic moment

The first measurement of magnetic moment of the neutrons was carried out by Alverez and Bloch (1940) using magnetic resonance technique. Cohen et al (1956) using a refined Rabi-method obtained a more accurate value for the neutron magnetic moment as \(\mu_n = -0.685039 \mu_p\) (\(\mu_p\) is the magnetic moment of the proton). The most accurate experimental value is due to Green et al (1979) obtained at ILL. This experiment has incorporated several innovations and improvements like use of cold neutrons (which allowed narrower resonance line widths), guide tubes and Ramsey separated oscillatory field magnetic resonance technique. The value for the neutron magnetic moment is
\[ \mu_n = -0.68497935 \mu_p. \]

### 3.5 The neutron lifetime

One of the important properties of free neutron is its instability. Since the neutron decay is the most elementary example of the beta decay, measurements of neutron half-life are of special importance, particularly in the determination of weak coupling constants \(g_v\) and \(g_A\). Neutron lifetime also plays a crucial role in astrophysics where it is central to the problem of helium production in the early universe (Taylor 1979). The neutron decays by weak interaction as
\[ n \rightarrow p + e + \bar{\nu} \]
according to the radioactive law with a decay rate
\[ \frac{dN(t)}{dt} = \lambda N. \]

In order to measure the half-life one has to determine the absolute number of neutrons in a well defined volume and obtain \(dN(t)/dt\) from the measured rate of decay particles either protons or electrons from this volume. The neutron density within the beam is obtained with a high degree of precision using foil (1/\(\nu\) detector) activation technique. The most formidable problem in any neutron lifetime experiment is to be able to detect the protons or electrons from the decayed neutrons in the presence of high gamma background.

Robson (1950) made the first serious attempt to measure its half-life accurately. He was able to detect both the decay particles, the proton and electron eliminating the background. The major uncertainty was the effective volume from which the decayed
particles emerged. He obtained the neutron half-life as $12.8 \pm 2.5$ min. Sosnovsky et al (1959) carried out precise measurements and came out with a neutron half-life as $11.7 \pm 0.3$ min. The drawback in these experiments was that only a small fraction of the decays occurring in the source was recorded directly since the solid angle for the collection was small. Christensen et al (1972) using an $8 \text{ kG}$ magnetic field were able to record all the decays. The magnetic field forces the decayed electrons leaving the beam to spiral about the field lines so that they eventually hit a large plastic scintillator. This measurement gave the half-life value as $10.61 \pm 0.16$ min. Byrne et al (1980) using "very cold neutrons" obtained $10.82 \pm 0.16$ min for the neutron half-life. The important features of this experiment were (1) a well-defined volume for the neutron beam (2) practically zero background, achieved by trapping the emerging protons for a significant time in an electromagnetic potential well before being released and detected. Figure 2 shows schematic of the experimental layout. Table 1 gives the basic properties of the free neutrons.

![Schematic of the experimental layout.](image)

**Figure 2.** Experimental layout for the determination of neutron lifetime.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass ($m_n$)</td>
<td>$1.008665012 \ (0.011 \text{ ppm})$</td>
</tr>
<tr>
<td>Charge ($q_e$)</td>
<td>$\leq (-1.5 \pm 2.2) \times 10^{-29} \text{ e}$</td>
</tr>
<tr>
<td>Spin $J$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>Magnetic moment ($\mu_n$)</td>
<td>$-0.68497935 \ (0.25 \text{ ppm})$</td>
</tr>
<tr>
<td>Half-life (weighted average)</td>
<td>$(10.68 \pm 0.13) \text{ min}$</td>
</tr>
<tr>
<td>Ratio of weak interaction</td>
<td></td>
</tr>
<tr>
<td>coupling constants ($g_A/g_v$)</td>
<td>$-1.255 \pm 0.006$</td>
</tr>
</tbody>
</table>

**Quantized properties**

- Intrinsic parity: $+1$
- Isospin: $\frac{1}{2}$
- Baryon number: $1$
- Lepton number: $0$
- Strangeness: $0$
- Charm: $0$
- Bottomness: $0$
- Topness: $0$
4. Neutron sources

Free neutrons do not occur in nature because of their short lifetime and therefore must be produced artificially. Neutrons form a part of the atomic nuclei of all elements (except hydrogen) where they are held together by short range nuclear forces. The neutrons can therefore be separated from the nuclei if the nucleus is somehow brought into an excited state whose energy exceeds the binding energy of the last neutron. Excitation of nuclei of selected isotope is generally accomplished by bombarding them with $\alpha$ particles, protons, deuterons or gamma rays. Table 2 gives binding energy of the last neutron for some important light nuclei.

4.1 Neutron production by charged particle reactions and $(\gamma, n)$ reactions

Excited compound nuclei after the capture of charged particles can emit neutrons. The cross-sections for the reaction depend on the energy of the bombarding particle and on the value of the electrostatic barrier of the nucleus. The cross-section increases as the energy of the charged particle increases and reaches a maximum value equal close to the geometrical cross-section of the target nucleus.

4.1a $(\alpha, n)$ reactions: This reaction has historical importance because it was used in the discovery of the neutron. The reaction in general can be written as

$$\alpha + zX^A \rightarrow z+2X^{A+2} + n + Q.$$ 

The $Q$ value is positive for exothermic reaction and negative for endothermic reaction. Some well-known examples are

$$\alpha + \text{Be}^9 \rightarrow \text{C}^{12} + n + 5.704 \text{ MeV},$$

$$\alpha + \text{B}^{11} \rightarrow \text{N}^{14} + n + 0.158 \text{ MeV},$$

$$\alpha + \text{Li}^7 \rightarrow \text{B}^{10} + n - 2.79 \text{ MeV}.$$ 

Radioactive $(\alpha, n)$ sources

These were the first amongst the neutron sources and even today they are used widely. Polonium-beryllium, radium-beryllium, plutonium-beryllium and americium-beryllium

<table>
<thead>
<tr>
<th>Nucleus</th>
<th>Binding energy (MeV)</th>
<th>Nucleus</th>
<th>Binding energy (MeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$^2$</td>
<td>2.225</td>
<td>B$^{10}$</td>
<td>8.440</td>
</tr>
<tr>
<td>H$^3$</td>
<td>6.258</td>
<td>B$^{11}$</td>
<td>11.456</td>
</tr>
<tr>
<td>He$^3$</td>
<td>7.719</td>
<td>C$^{12}$</td>
<td>18.720</td>
</tr>
<tr>
<td>He$^4$</td>
<td>20.577</td>
<td>C$^{13}$</td>
<td>4.937</td>
</tr>
<tr>
<td>Li$^6$</td>
<td>5.663</td>
<td>C$^{14}$</td>
<td>8.176</td>
</tr>
<tr>
<td>Li$^7$</td>
<td>7.253</td>
<td>N$^{14}$</td>
<td>10.553</td>
</tr>
<tr>
<td>Be$^9$</td>
<td>1.665</td>
<td>N$^{15}$</td>
<td>10.834</td>
</tr>
<tr>
<td>Be$^{10}$</td>
<td>6.814</td>
<td>N$^{16}$</td>
<td>2.500</td>
</tr>
</tbody>
</table>

Table 2. Binding energy of the last neutron in light nuclei (Beckurts and Wirtz 1964).
are the important radioactive (α, n) sources. Table 3 gives characteristics of some α radioactive and spontaneous fission isotopes. Table 4 gives maximum specific output of neutrons for 1 g of radioactive isotope for some (α, n) sources.

4.1b (d, n) reactions: This is an extensively used reaction for the production of neutrons. The general equation for this reaction is

\[ d + zX^A \rightarrow z+1X^{A+1} + n + Q \]

Typical examples,

\[ d + H^2 \rightarrow He^3 + n + 3.26 \text{ MeV}, \]
\[ d + H^3 \rightarrow He^4 + n + 17.58 \text{ MeV}, \]
\[ d + Li^7 \rightarrow Be^8 + n + 15.02 \text{ MeV}, \]
\[ d + C^{12} \rightarrow N^{13} + n - 0.28 \text{ MeV}, \]
\[ d + Be^9 \rightarrow B^{10} + n + 4.36 \text{ MeV}. \]

Since the binding energy of the deuteron is small (~ 2.226 MeV/nucleon) the compound nucleus formed by the capture of deuteron is always found in a highly excited state and consequently almost all (d, n) reactions are exothermic. H\(^2\)(d, n) He\(^3\) reaction is widely utilised for the production of monoenergetic neutrons because of expected good neutron yields for deuteron energies even below 1 MeV. The H\(^3\) (d, n) He\(^4\) is the best
among \((d, n)\) reactions for generating monoenergetic neutrons with high positive \(Q\) value.

4.1c \((p, n)\) reactions: Like \((d, n)\) reaction, this reaction has large neutron output compared to \((\alpha, n)\) reaction due to lower threshold energies. This reaction is extremely popular to produce monoenergetic neutrons with small proton energies. The general reaction is

\[
p + Z \, X^A \rightarrow Z + 1 \, X^A + n + Q.
\]

Generally lithium and tritium are used as targets for this reaction

\[
p + Li^7 \rightarrow Be^7 + n - 1.646 \text{ MeV},
\]

\[
p + H^3 \rightarrow He^3 + n - 0.764 \text{ MeV}.
\]

Assuming that this reaction is essentially due to \(\beta\) decay the \(Q\) value of the \((p, n)\) reaction is

\[
Q = E_\beta - Q_n,
\]

where \(Q_n = 0.762 \text{ MeV}\) is the \(Q\) value of \(\beta\) decay of the neutron and \(E_\beta\) is the maximum energy of the \(\beta\). Thus all \((p, n)\) reactions are endothermic if the target nuclei are stable.

4.1d \((\gamma, n)\) reactions: The emission of neutrons by the interaction of gamma rays with nucleus is known as nuclear photo effect. With increasing energy more complex reactions such as \((\gamma, 2n)\) \((\gamma, p)\) and \((\gamma, pn)\) etc are possible. The reaction equation is

\[
\gamma + Z \, X^A \rightarrow Z \, X^{A-1} + n + Q.
\]

The minimum \(\gamma\) energy required for this reaction to take place is equal to the binding energy of the neutron in the nucleus. The neutrons are monoenergetic if the \(\gamma\)-ray is monoenergetic. The threshold energy ranges from 10 to 20 MeV. The energy of the neutron can be calculated fairly accurately using the relation

\[
E_{\text{neutron}} = M_R \frac{(E_\gamma - Q)}{(M_R + m)},
\]

where \(M_R\) and \(m\) are the masses of target nucleus and the neutron respectively.

4.1e Radioactive \((\gamma, n)\) sources: The \(\gamma\) energy of most of radioactive substances rarely exceeds 3 MeV and therefore \((\gamma, n)\) reactions are possible only in beryllium \((Q = -1.665 \text{ MeV})\) and deuterium \((Q = -2.225 \text{ MeV})\). The drawbacks of these sources are their small yields and short half-life of the emitters. The target should be several centimeters thick to have good neutron output since cross-section for \((\gamma, n)\) is small. Sb\(^{124}\)Be neutron source is most widely used in practice due to the comparatively large half-life (60-9d) of antimony.

4.2 Neutrons from charged particle accelerators

Accelerator-based neutron sources have distinct advantage in that the neutron source strength and its energy can be controlled with ease. Further, the source can either be continuous or pulsed.

Acceleration of charged particles like protons, deuterons and \(\alpha\)'s is based on the interaction of their charge with an electric field. Essentially an accelerator is a device in
which the charged particles acquire the required energy by passing through a number of small potential differences whose sum is equal to the required $V$.

Accelerators are classified into different categories such as electrostatic, high frequency and induction accelerators depending on the way the electric field is generated to accelerate charged particles. Van de Graaff and Cockroft-Walton cascade generators belong to the first type where the charged particles acquire energy determined by the difference in potential between the electrodes.

Most important accelerators like cyclotron, microtron, synchrotron, synchrocyclotron and linear accelerators belong to the second group. In these accelerators, the particles interact with the high frequency electric field or cavity resonator through which they pass. Finally in the last group, the particles are accelerated under the action of an eddy electric field produced due to electromagnetic induction. Betatron is an example of this type of accelerator.

**Van de Graaff generator:** This is mainly used to accelerate protons and deuterons up to few MeV for production of monoenergetic neutrons using $(p, n)$ and $(d, n)$ reactions. Typical beam current is about $10 \mu$A from continuous mode operation and is about $10$ mA in the pulsed mode operation.

**Cascade generator:** This is similar to Van de Graaff except it differs in generating the high voltage. These are mostly used for generating small voltages to produce neutrons from $H^2(d, n)He^3$ or $H^3(d, n)He^4$ reactions.

**Cyclotron:** Protons and $\alpha$ particles can be accelerated to hundreds of MeV in cyclotrons. These energetic particles are used to produce neutrons using $(p, n)$ or $(\alpha, n)$ reactions.

**Other accelerators:** Synchrocyclotron, synchrotron isochronous cyclotron (azimuthal variation of field) and linear accelerators are basically meant for getting very high energy particles.

**Electron accelerator:** These accelerators can be used to accelerate electrons up to energies equal to 50–100 MeV. On striking a target, these electrons give intense bremsstrahlung with a continuous spectrum. The bremsstrahlung in turn produces neutrons by $(\gamma, n)$ processes. Bombarding a thick uranium target with 40 MeV electrons of about 1 mA current yields $10^{14}$ n/sec.

### 4.3 Neutrons from research reactors

Research reactors operating the world over are the main sources of copious supply of neutrons. The built-in facilities in the reactor enable carrying out a number of experiments simultaneous with the production of radioisotopes. Typically a 10 MW thermal reactor produces about $7.5 \times 10^{17}$ neutrons per second. Unfortunately only a small fraction of this source is actually available to the experimenter. Hence the continued endeavour is to evolve and optimize core designs that would maximize the neutron flux $\phi$ which is defined as the number of neutrons passing through unit area in all directions in each second. The physics aspects of the reactor are basically characterized by this flux.

Since the neutrons in the reactor have a broad distribution of energy, from 0 eV to 10 MeV, the flux is a function of energy $E$ of the neutron. Thus $\phi(E)\text{d}E$ is the flux of the neutrons. Three regions of energies are identified within the spectrum.

(a) **Thermal neutrons** ($E < 0.2$ eV). These neutrons are in thermal equilibrium with
the moderator atoms and their energy distribution follows a Maxwellian form given by

$$\phi(E) dE = \frac{2\pi n}{(\pi kT)^{3/2}} \left( \frac{2}{m} \right)^{1/2} E \exp(-E/kT) dE,$$

where \( n \) is neutron density, \( k \) is Boltzmann's constant, \( m \) is neutron mass and \( T \), the temperature of the medium in degrees Kelvin.

(b) Resonance neutrons: \((0.2 \text{ eV} < E < 0.5 \text{ MeV})\). In this range, the neutron spectrum is determined by the neutrons being slowed down by elastic collisions with the moderator nuclei. In this region the spectrum follows the behaviour

$$\phi(E) dE = \frac{\text{(constant)} dE}{E}$$

(c) Fast neutrons: \((E > 0.5 \text{ MeV})\). These neutrons have more or less the same energy distribution as that of fission neutron spectrum which is given by

$$N(E) \approx \exp(-E) \sin h(2E)^{1/2} \quad (E \text{ in MeV}).$$

The intensity of each of these components depends upon the type of reactor, its operating power and the point inside the reactor. The major limiting factor to have thermal fluxes better than \(10^{15} \) neutrons/sec/cm\(^2\) is the power density which is controlled by various requirements of heat transfer, heat transport and stability of core materials under irradiation. Two representative types of reactors will now be considered.

Research reactors fuelled with natural uranium and moderated with heavy water are in operation in different countries. Indian reactor cirus, Canadian reactors NRX and NRU and French reactor EL2 are examples of this type of reactors.

The second type of reactors are fuelled with highly enriched uranium and moderated with light water. Material testing reactor of US and other swimming pool reactors are examples of this category.

Table 5 gives the different components of the flux at the core centre for these reactor types operating at 10 MW.

A simple expression (Spinrad 1966) can be derived which relates the core average flux to the core parameters like critical mass, core volume and power density to select the type of reactor to have increased fluxes.

The core average flux \( \phi \), is related to the power density \( P_d (\text{MW/l}) \) and the macroscopic fission cross section \( \Sigma_f \) as

$$\phi = \frac{P_d \times 3.3 \times 10^{13}}{\Sigma_f}.$$
The $\Sigma_f$ is further related to critical masses and the core volume $V$ as

$$
\Sigma_f = \frac{N_0 \sigma_f}{A} (M/V),
$$

where $N_0$ is Avagadro's number, $\sigma_f$ is the microscopic fission cross-section and $A$ is the molar mass of the fuel. Substituting this in the first equation

$$
\phi = \frac{(3.3 \times 10^{13}) A P}{N_0 \sigma_f M},
$$

where $P$ is total power. Thus it is clear from these equations that to get high flux, one should attempt to have a highly dilute core ($M/V$ is small) or a small critical mass depending on whether power density or power is the limiting factor.

The beam tubes

Beam tubes are provided to make neutron beams available outside the reactor for carrying out the experiments. These tubes are normally cylindrical aluminium or zircalloy pipes which penetrate the shield in the horizontal direction to reach either the core or reflector or any other location to satisfy the experimental requirement. The neutron current density that can be obtained from such a tube can be calculated as follows. If $A$ is the front surface area of the beam hole and $\phi$ is the flux at that surface, then the number of neutrons entering the beam per second is $\phi A/4$. If $l$ is the length of the beam tube, then the neutron current density $J$ at the tube exist is

$$
J = \phi A/4\pi l^2.
$$

Beam tubes are normally positioned tangential to the core to reduce the gamma ray background. Pure thermal beams are extracted from the thermal column. For getting

![Figure 3. Schematic layout of the various neutron beam tubes in the reactor Dhruva.](image-url)
epithermal beams the tangential tubes are located nearer the core and cadmium filters may be used to suppress the thermal neutrons present. Gamma ray intensity can also be reduced with lead filters. Figure 3 shows the various beam tube locations and the proposed experiments in the reactor *Dhruva* which is undergoing commissioning tests at BARC.

4.4 Spallation neutron sources

As early as in 1947 E O Lawrence conceived the idea of using energetic charged particles (from accelerators) in conjunction with a target to produce neutrons. The next important milestone in this direction was a detailed design study by Wilson *et al* (1965) of AECL, Canada, of an intense neutron generator (ING) in the early 1960s. Subsequently, different laboratories including BNL, ORNL and LASL have undertaken programmes in this challenging field.

For all practical purposes spallation reactions are inelastic nuclear reactions in which one of the two collision partners is a complex nucleus (target) and the energy available is several times the interaction energy between the nucleons in the nucleus. Thus, if the incident energy exceeds something like 50–100 MeV per nucleon, it is referred to as a spallation reaction. Spallation reactions are described in terms of the two-step model suggested by Serber (1947). In the first step the incident high energy particle enters the target nucleus and interacts with some of the nucleons in the target in what is known as an intranuclear cascade. In this process few nucleons will be knocked out of the nucleus. The excited residual nucleus deexcites in the second step of this model emitting a number of single nucleons. The first stage is known as 'fast' stage, knock on stage or cascade stage. Similarly the second stage is referred to as slow stage, deexcitation stage or evaporation stage.

The yield of such spallation neutrons is a function of: (a) incident particle type and energy (b) target material and its size and geometry. Figure 4 shows some results of experiments conducted with the protons from the 3 GeV cosmotron at BNL. It can be seen that the neutron yield from U targets is twice that from lead target. This is simply due to fast fission effect in $^{238}$U.

The neutron yield calculations have been carried out by Monte-Carlo code called HETC (high energy transport calculations) (Chandler 1972) developed at ORNL. This

![Figure 4. Neutron yields obtained by bombardment of heavy targets with high-energy protons.](image-url)
code is a high energy version of an earlier code designated nmtc (nuclear meson transport code) (Coleman 1970). Detailed system analysis carried out at BNL and AECL indicate that: (i) the optimum proton energy is likely to be between 800 MeV and 1 GeV (ii) the optimum beam current is about 300 mA.

Linear accelerators capable of accelerating a steady continuous beam of protons at a current of ~ 100 mA are available today. Synchrotron is another potential candidate for getting high energy charged particle beams. Table 6 gives a number of spallation neutron source facilities.

4.5 Dense plasma focus as a neutron source

The dense plasma focus (Mather 1964; Fillipov 1962; Imshennik 1973) is an intense source of monoenergetic neutrons. Essentially this is a fast dynamic Z-pinch in which stored magnetic energy is rapidly converted into plasma energy and then compressed by its self-magnetic field. The device shown in figure 5 consists of two cylindrical electrodes in a coaxial configuration and a capacitor bank. The space between the electrodes is filled with deuterium gas at a pressure of few torr. The electrodes are connected to a high energy (1 kJ to 1 MJ), high voltage (20 to 80 kV), low inductance (~ 40 nH) capacitor bank through a spark gap. As soon as the spark gap is closed, the capacitor bank is discharged and the gas breaks down at the insulator, allowing a radial

Table 6. Spallation neutron source facilities.

<table>
<thead>
<tr>
<th>Name</th>
<th>LAMPF</th>
<th>SIN</th>
<th>TRIUMF</th>
<th>ING</th>
<th>RUTH. LAB.</th>
<th>IPNS-1</th>
<th>KEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>U.S.A.</td>
<td>Swiss</td>
<td>Canada</td>
<td>Canada</td>
<td>U.K.</td>
<td>U.S.A.</td>
<td>Japan</td>
</tr>
<tr>
<td>accr.</td>
<td>Linear</td>
<td>S.O.C.</td>
<td>Cyclotron</td>
<td>Linear</td>
<td>Sync</td>
<td>His</td>
<td>Sync</td>
</tr>
<tr>
<td>P energy</td>
<td>800 MeV</td>
<td>590 MeV</td>
<td>500 MeV</td>
<td>1 GeV</td>
<td>800 MeV</td>
<td>600 MeV</td>
<td>500 MeV</td>
</tr>
<tr>
<td>Avg. current</td>
<td>1 mA</td>
<td>100 µA</td>
<td>100 µA</td>
<td>65 mA</td>
<td>200 µA</td>
<td>—</td>
<td>1.5 µA</td>
</tr>
<tr>
<td>Target</td>
<td>W Pb, Bi</td>
<td>Molten lead</td>
<td>Pb-Bi</td>
<td>DEP, U</td>
<td>238U DEPT. U</td>
<td>DEPT. U</td>
<td></td>
</tr>
<tr>
<td>N. yield/s</td>
<td>$2 \times 10^{16}$ n/sec</td>
<td>—</td>
<td>$10^{16}$ n/sec</td>
<td>$10^{16}$ n/sec</td>
<td>$4 \times 10^{16}$ n/sec</td>
<td>$2 \times 10^{16}$ n/sec</td>
<td>$1.8 \times 10^{14}$</td>
</tr>
<tr>
<td>Pulse rate</td>
<td>120</td>
<td>Con/pulse</td>
<td>—</td>
<td>C.W.</td>
<td>50</td>
<td>45</td>
<td>20</td>
</tr>
<tr>
<td>Power cons</td>
<td>27 MW</td>
<td>68 MW</td>
<td>56 MW</td>
<td>100 MW</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Target heat</td>
<td>200 kW</td>
<td>45 kW</td>
<td>38 kW</td>
<td>38 MW</td>
<td>450 kW</td>
<td>20 kW</td>
<td>2 kW</td>
</tr>
</tbody>
</table>

Figure 5. Schematic of a plasma focus discharge showing initial and final phases.
current to flow between the electrodes. This radial current produces an azimuthal magnetic field. The combination of the azimuthal field with a radial current produces $J \times B$ force on the ions away from the insulator. Thus as the current continues to flow the plasma front is accelerated along the $z$ axis, ionizing the neutral gas it encounters. As soon as the plasma front reaches the end of the electrode, the axial momentum of the plasma causes the front to fold over the end and radially imploding towards the axis of central electrode creating a hot dense plasma focus typically about 2 mm in dia and 10 to 30 mm long. The discharge current at the time of focus formation can be as high as several hundred kiloamperes to a few mega amperes. Intense bursts of electrons, neutrons, ions, soft x-rays ($<10$ keV) and hard x-rays ($>30$ keV) are observed to emanate from the focus.

A unique feature of plasma devices is, irrespective of differences in design, the neutron yield $N$ of all optimised plasma focus devices is found to obey a universal scaling law, namely that the yield is found to scale as the fourth power of peak discharge current ($I$) or as the square of the capacitor bank energy ($E$). Neutron yield (Conrad 1983) as function of plasma current for different focuses of various international laboratories including BARC (Shyam and Srinivasan 1978) is shown in figure 6. During the 1970s plasma focus devices have produced about $10^{12}$ neutrons per discharge from (D–D) reactions. With the present state of art $10^{14}$ neutrons per discharge can be achieved by dissipating about 300 kJ of stored energy.

4.6 Classification of neutrons on the basis of energy

Neutrons are classified as ultra cold neutron (UCN), very cold neutrons (VCN), cold neutrons (CN), thermal neutrons, epithermal neutrons and fast neutrons depending on their energy (or wavelength). Since each class of neutrons has a role to play in the basic research, one has to tailor the easily available neutrons of a given energy. Of all the

![Figure 6. Neutron yield as a function of plasma current for different foci.](image-url)
neutron sources \text{ucn} and \text{vcn} beams are difficult to get with good intensities.

A very useful relation between neutron velocity $V$ (m/sec), de Broglie wavelength ($\lambda$) and the neutron energy $E$ (eV) is given by

$$E \ (\text{eV}) = 5.2 \times 10^{-9} \ V^2 \ (\text{m/sec}),$$

$$= 8.2 \times 10^{-2} \ \lambda^{-2} \ (\text{A}).$$

\textbf{Ultra cold neutrons:} Neutrons entering a solid material from vacuum with energy $E$ less than $U_{\text{eff}}$ (called the locally averaged Fermi scattering potential for the material), with a corresponding velocity less than $v_c$ or wavelength greater than $\lambda_c$ will be reflected even at normal incidence. Such neutrons are called ultra cold neutrons (\text{UCN}). Typically for \text{UCN} $E \leq 2.4 \times 10^{-7}$ eV, $v \leq 6.8$ m/sec and $\lambda \geq 580$ A.

\textbf{Very Cold Neutrons:} Neutrons with the velocity $6.8$ m/sec $\leq v \leq 100$ m/sec are referred to as very cold neutrons. These neutrons are reflected at angles $< 90^\circ$. The corresponding energy and wavelengths are

$$2.4 \times 10^{-7} \text{ eV} \leq E \leq 5.2 \times 10^{-4} \text{ eV},$$

$$580 \text{ A} \geq \lambda \geq 126 \text{ A}.$$

\textbf{Cold neutrons:} These neutrons are reflected at very small angles of incidence and have velocities around $800$ m/sec with a corresponding energy of $0.0033$ eV and a wavelength of $5$ A.

Although \text{ucn}'s, \text{vcn}'s and \text{cn}'s exist as the low energy fraction of the Maxwell-Boltzmann distribution, it is almost impossible to extract them directly. So it is normal practice to generate these sources from a fast source by down scattering in a good converter or cold source at appropriate temperature.

\textbf{Thermal neutrons:} Neutrons which are in thermal equilibrium with the surrounding medium and follow Maxwellian distribution for neutron density with energy.

\textbf{Epithermal or resonance neutrons:} Neutrons in the energy range from about $0.2$ eV to $500$ keV belong to this category. The spectrum is determined by the neutrons being slowed down by elastic collisions with the moderator nuclei.

\textbf{Fast neutrons:} Neutrons with energy from $500$ keV to $20$ MeV.

\textbf{Super fast neutrons:} Neutrons possessing energy greater than $20$ MeV.

5. Applications

The important properties of the neutron are its mass, charge neutrality, spin, magnetic moment and lifetime against $\beta$ decay etc. These unique properties have made the neutron an ideal tool for the scientist to probe the unknown mysteries of the nature. The ever expanding research areas where the neutron is playing a dominating role encompasses the physical sciences, biological sciences, medical sciences, earth sciences, agricultural sciences not to mention its important role in nuclear power production and other industrial applications (Schofield 1982).

Exhaustive research work over the years on structural studies and dynamical properties of condensed matter using neutrons has been pursued at \text{BARC} (Iyengar 1973). Study of properties of magnetic materials using the inherent magnetic moment
of the neutron (Satya Murthy and Madhav Rao 1981), production and applications of polarised neutrons (Satya Murthy and Madhav Rao 1984) will also be not discussed here.

One of the most outstanding applications of the neutron is in power production through fission chain reactions in a reactor. The behaviour of the neutron population in such a reactor can be obtained by solving the Boltzmann neutron transport equation in seven variables. The input to solve the transport equation is the neutron cross-section data. Ramanna et al (1956) developed the pulsed neutron technique in our country to generate thermal cross-section data for moderators and coolants. The impact of nuclear research on the future technology of nuclear power was discussed by Iyengar (1979). In view of the availability of excellent text books (Bell and Glasstone 1970) we will not discuss this topic also.

5.1 Sub-Lawson fusion systems for power production and breeding

In all likelihood, the first generation commercial fusion reactors are to be fuelled with a mixture of deuterium (D) and tritium (T) (Hirsch 1975; Ramanna 1984) wherein about 14* 1 MeV out of 15 MeV energy released in the fusion reaction will be carried away by the neutrons liberated in the reactions. The tokamak and the tandem mirror are the top contenders in the field (Ramanna 1984). The plasma zone of any such device is surrounded by a medium, known as blanket, wherein the 14 MeV neutrons leaking out of the plasma zone deposit their energy for the subsequent generation of power. A conceptual design of a fusion reactor based on a magnetic confinement is shown in figure 7. For a pure fusion system to be economical, it is expected that the power amplification factor $Q$ (fusion output power/power in the plasma) should be of the order of 10 (Dolan 1982). Such high $Q$ values are unlikely to be achieved in the near future without further technological breakthroughs.

The hybrid concept (Lidsky 1975) in which fusion and fission systems are coupled exploits the fact that fusion reactions are neutron rich but energy poor while fissions are energy rich but neutron poor. The two hybrid concepts that are vigourously pursued are (Maniscalce 1981): (a) hybrid blankets or fast fission blankets and (b) symbiotic blankets or suppressed-fission blankets. These two concepts are shown schematically in figure 8.

Figure 7. Conceptual design of a fusion reactor based on magnetic confinement.
In the fast fission blanket system, the D-T fusion source is surrounded by a fertile material blanket. The fusion neutrons cause fast fissions in the fertile material, thereby amplifying the fusion energy and also multiplying the fusion neutrons. The fission energy will be utilized to generate power just like in any nuclear power reactor. One of the multiplied neutrons is needed to breed tritium from lithium situated in the blanket and the remaining are available for breeding fissile fuel from fertile materials like $^{232}$Th and $^{238}$U.

Suppressed fission blanket concept is essentially meant for producing fissile fuel from fertile materials. An additional non-fission neutron multiplying zone such as BeO, lead etc is placed between the D-T fusion source and the fertile blanket (Nargundkar et al 1984). The neutron multiplier also moderates the 14 MeV fusion neutrons to below fission threshold energy (1.5 MeV) before they reach the fertile zone, thus suppressing fissions in the blanket while breeding the fissile material.

Although the basic motivation for developing a fusion breeder is essentially the same as for the fission breeder i.e. to exploit the enormous energy potential of the world’s
abundant Th-232 and U-238 natural resources, there are important differences that make the fusion breeder a better choice. The following advantages are noteworthy:

(i) Fusion breeder needs no fissile inventory. (ii) Power densities are 10 to 200 times less than those in the fission breeder. (iii) The fuel production and energy production can be separated. (iv) Fusion system can produce many times (~30) more net fissile product per unit power than a fast breeder. (v) Since enormous amounts of fissile fuel can be produced from fusion breeder, many fission reactor concepts can be adopted, not necessarily the fast breeders. (vi) Fusion neutrons can also be used to transmute the longlived radioactive fission waste products to less hazardous and toxic substances with reduced half-lives. This solves the vexing problem faced by the fission reactors. (vii) There are no safety problems since power is generated in a subcritical system.

5.2 Neutron radiography

Neutron radiography, a relatively new and powerful technique of producing images of objects, is finding applications in varied and diverse fields such as engineering, metallurgy, reactor technology, aerospace, medicine, ordnance and electronics etc. Hitherto the most established technique of flaw detection in industrial products was through x-ray radiography. However this method has limitations and fails to give required information under certain situations like locating light material behind or within heavy materials, testing of radioactive parts or irradiated fuel assemblies of power reactors, detection of hydrogenous inclusions in materials and inspection of thick plates of heavy materials. In all these situations neutron radiography can be used successfully because the way neutrons interact with matter is totally different from the way x-rays interact. X-rays interact with the orbital electrons only and hence the attenuation coefficient increases with increasing Z unlike the neutrons which interact with the nucleus of the atom. Another important difference is the ability of the neutrons to distinguish between the isotopes of same element which is impossible with x-rays.

The technique of neutron radiography involves the placement of the sample under test in the path of a well-collimated beam of thermal neutrons and studying the transmitted beam modified by the attenuations within the sample. Since the photosensitivity of x-ray films to neutrons is very low, the transmitted neutron beam is made to fall on a suitable converter screen which in turn emits photographically sensitive radiations. These radiations finally expose a conventional x-ray film to give the required radiograph.

In the first method, the converter screen along with the x-ray film is placed behind the object under test and exposed sufficiently long to get the image. The converter screens are thin sheets of Rh, Gd, Cd, Sm or a suitable scintillator loaded with Li-6 or B-10. This method has some drawbacks if gamma ray background is present in the neutron beam or the specimen itself is radioactive. The second technique called the transfer method circumvents this problem. In this, a converter screen alone is exposed to the transmitted neutron beam and the screen becomes radioactive. The latent image of the specimen formed on the screen is then transferred to the photographic film in a separate place free from any background. Suitable converter materials, keeping the half-life in mind are In, Dy and Au.

The third technique of recent origin uses Li-6, B-10 or U-235 loaded screens to convert the thermal neutrons into alpha particles or fissions fragments. These charged particles fall on a special plastic material causing radiation damage along their paths.
The radiograph is obtained after etching the plastic sheet in a chemical solution. The advantages of this method are (i) totally gamma insensitive, (ii) no limitation on time of exposure and (iii) elimination of use of x-ray films.

In addition to what has been mentioned some specific applications in nuclear technology are: (a) testing of unirradiated nuclear fuel assembly to give qualitative information on assembly, voids, foreign materials and quantitative information on homogeneity, average density etc. (b) testing of irradiated fuel to give information on leaks, densification, swelling, migration and disintegration etc.

A dedicated neutron source reactor specially meant for neutron radiography applications is being set up at the Reactor Research Centre, Kalpakkam (Srinivasan et al 1979) for testing of fuel assemblies of fast breeder test reactor and for other related applications.

6. Conclusion

Neutrons by virtue of their unique and unparalleled properties have provided scientists and engineers with a tool to unravel the mysteries of nature. Every important property of the neutron has been exploited by the researchers to help probe into the fundamental properties right from understanding the nature underlying the nuclear forces to the formation and the structure of the stars. With further breakthroughs in technology more and more intense neutron sources tailored to specific needs will become increasingly available that would boost the neutron utility further both in the fields of basic and applied sciences. Neutrons have already contributed in a large measure to global human welfare and would undoubtedly continue to do so in an ever increasing manner in years to come. It is no prophecy therefore to predict that the neutron which celebrated its golden jubilee in the year 1982 will have many more jubilees in store in the future.

References

Alvarez L W and Bloch F 1940 Phys. Rev. 67 111
Bothe W and Becker G H 1930 Z. Phys. 66 289
Bell I B and Glasstone S 1970 Nuclear reactor theory (New York: Van Nostrand)
Chandler K C 1972 Report ORNL 4744
Coleman W A 1970 Report ORNL 4606
Conrad H 1983 Nuclear technologies in a sustainable energy system (eds) G S Baner and A McDonald (New York: Springer-Verlag)
Fillipov V V 1962 Nucl. Fusion Suppl. 2 577
Halpern T et al 1937 Phys. Rev. 52 142
Imshennik V S 1973 *Nucl. Fusion* 13 929
Iyengar P K 1973 Sixtieth Session of the Indian Science Congress, Chandigarh
Iyengar P K 1979 15th Founder Memorial Lecture, Shri Ram Institute for Industrial Research, Delhi
Lidsky L M 1975 *Nucl. Fusion* 15 151
Maniscalce J A 1981 *Nucl. Technol./Fusion* 1 419
Mather J W 1964 *Phys. Fluids, Suppl.* 7 52
Nargundkar V R et al 1984 *Fusion Technol.* 6 93
Ramanna R et al 1956 *J. Nucl. Energy* 2 145
Ramanna R 1984 *Prospects of fusion power*, Meghnad Saha Memorial Lecture, SINP, Calcutta
Robson J M 1950 *Phys. Rev.* 78 311
Rutherford E 1920 *Proc. R. Soc. London* **A97** 373
Satya Murthy N S and Madhav Rao L (eds) 1981 *Current trends in magnetism* (Bombay: Indian Physics Association)
Satya Murthy N S and Madhav Rao L 1984 *Pramana* **22** 313
Schofield P 1982 *The neutron and its applications* Conf. Series No. 64 (London: Institute of Physics)
Schwinger J S 1937 *Phys. Rev.* **52** 1250
Serber R 1947 *Phys. Rev.* **22** 1114
Shaprio I S and Estulin I V 1956 *Sov. Phys. JETP* 3 626
Spinrad B I 1966 *Seminar on intense neutron sources* (eds) G R Keepin and H T Motz 319 Conf. 660925
Stedman R 1968 *J. Phys.* **1** 1168
Taylor R J 1979 *Nature (London)* **282** 559
Wilson et al 1965 AEC Symposium on the Generation on Intense Neutron Fluxes (ed. W G Lewis) AECL 2177
Risk evaluation and protection against ionizing radiation

V K JAIN
Health Physics Division, Bhabha Atomic Research Centre, Bombay 400085, India

Abstract. The advent of nuclear reactors ushered in an era of increasing number of sources of ionizing radiations. However, the potential of ionizing radiations to cause harmful effects was recognized soon after the discovery of x-rays and radioactivity i.e. long before the building of nuclear reactors. Therefore, protection against ionizing radiations has been of paramount concern and has guided the development of atomic energy and related fields. The advances in technology in general resulted in an increase in accidents causing injury and death. It was realised that even medicines, food additives and a host of other substances of daily use had injurious side effects. Smoking was found to be extremely harmful. From these emerged the concepts of quantitative and relative risks. This article discusses briefly the concept of risk vis-a-vis ionizing radiations and approaches to protection against them.

Keywords. Risk evaluation; ionizing radiation; radio therapy; cancer induction; somatic effects; diagnostic radiology; radiation protection.

PACS No. 28.80; 87.60

1. Introduction

X-rays and radioactive emanations, both were in common use soon after their discoveries. That these radiations might be harmful at high doses was also quickly realized. Voluntary individual and coordinated efforts towards protection were taken up early. Scientific bodies in countries like Germany (1913), UK (1915) and the USA (1916) were reviewing the requirement of their safe use. In 1925 the International Congress of Radiology was formed and under its patronage were born the International Commission on Radiation Units and Measurements (ICRU) in 1925 and the International Commission on Radiological Protection (ICRP) in 1928.

The ICRP issued its first recommendations in 1928 itself and the latest major revision came in 1977 (ICRP 1977a). While the presence of radiations and radioactivity has tremendously proliferated during the 90 years since their discovery considerable information and knowledge has also become available concerning their interactions with and the effects on the living systems. More is probably known about the consequences of exposure to ionizing radiations than any other environmental agent or occupational hazard. Paradoxically greater knowledge about the consequences of radiation exposure has led to its becoming an issue of serious concern to society (Pochin 1980; Goldman 1982). Partly this is due to the growing awareness of risks encountered in day-to-day experience: viz that medicines have side effects, food colours may be harmful, industrial effluents are hazardous, pesticides leave damaging residues and a host of similar other findings, besides the usual accidents, house collapses and fires, etc. Detailed studies of occupational mortality risks have been carried out with a view to have the right perspective of risks in different occupations (ICRP 1977b; Cohen 1981).
2. Evolution of the concept of risk

For a long time many scientists believed in a threshold level of radiation exposure below which no harm would occur and which was, therefore, thought to be a safe level. But data have been accumulating over the years which show that any exposure, however small, may not be considered so low that there was no risk due to it. It, therefore, became essential to make a quantitative assessment of the amount of risk or the degree of safety ensured at the prescribed level of dose limits. But this immediately leads to questions such as "what is the amount of risk which is acceptable", "how safe is safe enough?" (Fischhoff et al. 1979; Watson 1981; Darby and Keeney 1981). Answers to these questions involve not only technical but economic as well as social aspects. Equally important are individual and collective public responses to risk. Perception of risk is most likely to have variants such as (i) economic constraints (ii) bias (iii) strong personal beliefs and (iv) the format of presentation. It is, therefore, very important to place risks in proper perspective. This can be done by placing them in a comparative order with other hazards or different sources of the same hazard. Ionizing radiations, for example, are encountered as natural background, in the environment, medicinally, occupationally and so on. This approach has led to the realisation that radon gas emanating from construction materials and accumulating in closed or poorly ventilated buildings is a major source of public exposures. The acceptability of risk is associated with the benefits derived from the source of the risk. It also follows that the acceptable risk would be the risk associated with the best of the available alternatives.

There are various approaches to determining the best alternative, but four methods frequently employed are cost-benefit analysis, revealed preferences, expressed preferences and natural standards (ICRP 1982; Starr et al. 1976). The most used index of risk is the fatality rate both for natural disasters as well as human activities. It is observed that risks of the order of $10^{-3}$ fatalities per person per year are non-existent and are obviously in the category of unacceptable. Risks of the order of $10^{-6}$ per year, on the other hand, do not cause any anxiety to the average person and are hence acceptable. For risk levels between these limits, society spends money to reduce them, or pays compensation or derives commensurate benefits to put up with the inconvenience. Starr (1969, 1981) who studied these rates arrived at a societally acceptable and unacceptable benefit risk pattern of involuntary exposures shown in figure 1. He also analysed the

![Figure 1. Benefit-risk pattern of involuntary exposure (Source: Starr 1981).](image-url)
voluntary and involuntary aspects of risk, depicted in figure 2, and derived what might be called the "laws of acceptable risk" viz

(i) the acceptability of risk is roughly proportional to the third power of the benefits;
(ii) there seems to be a willingness on the part of the public to accept risks from voluntary activities, roughly a thousand times greater than it would tolerate from involuntary activities;
(iii) there is an inverse relationship between the acceptable level of risk and the number of persons exposed to that risk.

3. Ionizing radiations

3.1 Nature and measurement of ionizing radiations

The electromagnetic or particulate radiation interacts with matter causing physical disruption of the neutral atom into a knocked-out electron and the residual electron-deficient atom, the two constituting an ion pair. The process is called ionization. X-rays, gamma rays and beta particles produce these ion pairs in sparse paths and are called low LET (linear energy transfer) radiations. By contrast alpha particles, protons and heavy ions produce ion pairs in closely spaced paths and are called high LET radiations. The measurement of ion pairs produced in air is the first step in the measurement of radiation. The roentgen (R) is that amount of radiation which produces 1 esu of charge, positive or negative, in 1 c.c. of dry air at ntp. In the new international system of units (si) radiation exposure in air is measured in units of coulomb/kg. The production of ion pairs causes absorption of energy leading to absorbed radiation dose. The deposition of 100 ergs/g of matter constituted 1 rad of absorbed dose. The si system has replaced rad by Gray, symbol Gy, and equals 1 J/kg: 1 rad = 0.01 Gy.

The specification of absorbed dose is a physical event. The differences in the LET of different radiations causes differences in microscopically distributed energy in matter. To account for these differences, which ultimately lead to differences in biological effects an LET-dependent multiplier factor \( Q \), called the quality factor, is used to obtain a new quantity called the dose equivalent (DE). The factor \( Q \) has values ranging from 1 to
20. The dose is measured in units of rem or its SI equivalent Sievert, symbol Sv; 1 rem = 0.01 Sv (ICRU 1980; Jain and Soman 1978).

3.2 Radiation effects on the living system

Ionizing radiations are known to affect living cells by killing, interfering with their normal functioning like delay in cell division, inducing chromosomal aberrations and gene mutations, etc. However, the living system is also capable of self-repair of at least some types of insults to it. Therefore, damages to the living organism are intervened by the repair mechanisms particularly at low doses of low LET radiations. The high cellular sensitivity to radiation (radiosensitivity) is governed by certain characteristics of the cell system like (i) primitiveness (ii) proliferative activity (iii) differentiation (iv) number of divisions required to attain specialisation/maturity etc. Lymphocyte is an exception to these general observations. Table 1 lists cellular sensitivity in the decreasing order. However, potential radiation-induced cancer risk is not necessarily correlated with rapid cell division in tissues highlighting perhaps the complexity of cancer etiology.

3.3 Acute effects of radiation

Information on the effects of radiation has come either from laboratory experiments on animals or persons exposed accidentally to high doses of radiation. In vitro studies of single cell preparations show that cell survival and reproductive capacity can be diminished by relatively high radiation doses. Uniform irradiation of the human body leads to changes in the blood count picture, the extent of which depends on the dose. The lymphocyte being the most sensitive, its numbers decrease rapidly and remain so for long periods. Hemoglobin being the least sensitive is affected least and decreases slowly. Other components of the blood have a response which is between these two extremes and is shown in figure 3 for neutrophils and figure 4 for platelets (Paretzke et al. 1982). Table 2 gives the acute effects of whole body irradiation at different doses. Large total body doses impair most severely regenerative cells of the bone marrow, proliferative cells of the intestinal epithelium and the endothelial cells of the microvascular system. As a consequence a number of effects evolve in which phagocytic function is impaired, electrolyte balance is not maintained and gastrointestinal integrity is affected. The result is loss of defenses to infection and spread of internal infections.

<table>
<thead>
<tr>
<th>Table 1. Cell sensitivity to radiations in the decreasing order.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lymphocytes</td>
</tr>
<tr>
<td>2. White blood cells and red bone marrow</td>
</tr>
<tr>
<td>3. Intestinal crypt cells</td>
</tr>
<tr>
<td>4. Cells of the reproductive organs (gonads)</td>
</tr>
<tr>
<td>5. Liver cells</td>
</tr>
<tr>
<td>6. Cells of glands</td>
</tr>
<tr>
<td>7. Cells of the connective tissue</td>
</tr>
<tr>
<td>8. Kidney cells</td>
</tr>
<tr>
<td>9. Muscle cells</td>
</tr>
<tr>
<td>10. Nerve cells</td>
</tr>
</tbody>
</table>
Figure 3. Acute effect of irradiation on neutrophils (Source: Paretzke et al 1982).

Figure 4. Acute effect of irradiation on platelets (Source as in figure 3).
Table 2. Effects of acute whole-body radiation doses (based on Glasstone and Sesonske 1963).

<table>
<thead>
<tr>
<th>Acute dose, rems</th>
<th>Probable clinical effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 50†</td>
<td>No observable effects.</td>
</tr>
<tr>
<td>50† to 100</td>
<td>Slight blood changes but no other observable effects.</td>
</tr>
<tr>
<td>100 to 200</td>
<td>Vomiting in 5 to 50% of cases within 3 hours, with fatigue and loss of appetite. Moderate blood changes. Recovery will occur in all cases within a few weeks.</td>
</tr>
<tr>
<td>200 to 600</td>
<td>For doses of 300 rems and more, all exposed individuals will exhibit vomiting within 2 hours or less. Severe blood changes accompanied by hemorrhage and infection. Loss of hair after 2 weeks for doses over 300 rems. Recovery in 20 to 100% within 1 month to a year.</td>
</tr>
<tr>
<td>600 to 1000</td>
<td>Vomiting within 1 hour, severe blood changes, hemorrhage, infection, and loss of hair. From 80 to 100% of exposed individuals will succumb within 2 months; those who survive will be convalescent over a long period</td>
</tr>
</tbody>
</table>

† This number is sometimes given as 25. However, there is a recent tendency to believe that observable effects of radiation do not begin until about 50 rem.

Three different radiation syndromes, following acute exposure, have been noted. The first of these, occurring at 2 to 4 Gy, is the bone marrow syndrome in which due to heavy killing of blood cells, particularly stem cells of the bone marrow, there is very high susceptibility to infection and loss of nourishment. At somewhat higher doses, 4–6 Gy, the gastro-intestinal tract is severely affected leading to internal hemorrhage and infections. The central nervous system syndrome follows at very high doses of 10 Gy or more.

3.4 Low level radiation effects and the latent period

Low levels of radiation exposure do not cause any immediately observable effect. Protraction of dose over a long period of time causes much less severe effects than if the same dose were delivered acutely. Several animal studies have shown these differences in radiation effects. To cite one example a recent study on the effects of strontium 90, a low LET beta particle emitting radionuclide that concentrates in the skeleton shows that bone cancer is induced only when the radiation dose rates exceeded about 1 rad per day for several years i.e. after several hundred rems of dose (Raabe et al 1981). The American National Council on Radiation Protection and Measurements collected tumourigenesis data on some 12 different studies on laboratory animals covering doses upto 200 or 300 rem (ncrp 1980). They all show that low-dose rate radiation is about two to ten times less carcinogenic for the same doses compared with radiation delivered at high rates.

There is considerable incidence of natural cancer which in general increases with age. Each incremental dose of radiation causes an increase in this rate but with a time lag following irradiation in which no medical evidence is seen. This time lag is referred to as the latent period as explained in figure 5 (nas 1980). For most solid tumors this is of the order of 10–15 years and for leukemia it is about 3–5 years. The excess rate continues for a further period of about 25 years for leukemia and lifetime for solid cancers. This is known as the period of expression (nas 1980). Thus the principal, perhaps the sole,
radiation effect, at low-dose ranges, in the exposed population is the somatic risk of increase in probability of cancer induction. It is important that the evaluation of this increase in probability of cancer induction i.e. risk be based on human experience and not on animal studies.

4. Somatic effects: Cancer induction

The most important somatic effect of concern for radiation protection purposes is the induction of cancer in different tissues. Table 3 gives the sensitivity to natural and radiation-induced cancer of various sites. To determine this a considerable number of epidemiological surveys have been made of human populations who have been exposed to known absorbed doses and have been followed with adequate ascertainment of cancer incidence or mortality, for long periods that are needed to estimate the frequency of cancer induced by radiation (UNSCEAR 1977). Reliable information in this connection has come from three main types of exposures (i) radiological procedures, therapeutic (for benign conditions) and diagnostic, (ii) Japanese survivors of atom bomb attack and people of Marshall Islands affected by fallout from weapon testing, (iii) occupational-exposure in uranium and other hard rock miners and in radium dial painters.

4.1 Exposure in radiotherapy and diagnostic radiology

Ankylosing spondylitis has been treated with x-rays as well as intravenous injection of radium 224. The former led to enhanced mortality from leukemia related to bone marrow dose and the latter to an increased incidence of bone sarcoma (Court Brown and Doll 1965; Spiess and Mays 1973). Certain other organs such as stomach, lung lymphoid tissues and intestine exposed within the radiation fields also showed increased mortality due to cancer related to the mean absorbed dose in these organs (Smith and Doll 1978). Similarly other radiotherapeutic treatments have yielded
Table 3. Organ sensitivity to radiation-induced cancer

<table>
<thead>
<tr>
<th>Site of Cancer</th>
<th>Natural incidence</th>
<th>Radiation-induced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female breast</td>
<td>Very high</td>
<td>High</td>
</tr>
<tr>
<td>Leukemia</td>
<td>Moderate</td>
<td>Very high</td>
</tr>
<tr>
<td>Thyroid</td>
<td>Low</td>
<td>Very high</td>
</tr>
<tr>
<td>Lung</td>
<td>Very high</td>
<td>Moderate</td>
</tr>
<tr>
<td>Stomach and colon</td>
<td>High</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

estimates of cancer induction in breast, thyroid, brain, skin, etc (Shore et al 1977; Baral et al 1977; Hempelmann et al 1975; Modan et al 1974; Shore et al 1976). Treatment of benign diseases of the uterus by irradiation has provided risk estimates of pelvic organs (Smith and Doll 1976). Patients treated for thyroid cancer by radioiodine have given indications of leukemia incidence (Pochin 1969).

Diagnostic exposures being much lower, only in three situations the data has yielded information on cancer induction: firstly breast cancer increase in female patients who underwent repeated fluoroscopy during lung collapse therapy for tuberculosis (Boice et al 1979), secondly incidence of leukemia and liver cancer in patients who were given thorotrast as a contrast medium (UNSCEAR 1977); and thirdly increase in leukemia and other malignancies when the foetus got exposed during pelvic or abdominal radiographic examination of the mother (Stewart and Kneale 1970).

4.2 Japanese survivors and Marshall islanders

The exposed surviving populations from the cities of Hiroshima and Nagasaki (victims of atom bomb attacks) have shown enhanced frequencies of leukemia and at least nine other forms of malignancy namely the lung, female breast, thyroid and salivary glands, lymphoid tissues, bladder and urinary organs, aesophagus, stomach and large intestine (Beebe et al 1977). Although dose estimates are being revised (Lowe 1982) which may change the earlier derived risk factors, the fact remains that these studies have provided extensive human experience. Fallout from a weapon test on Bikini in 1954 resulted in ingestion of radio-iodine by some Marshall islanders. These have provided risk estimates for cancer of the thyroid (Conrad et al 1975).

4.3 Occupation exposures

Estimates of lung cancer induction by alpha radiation from radon daughter products to the bronchial epithelium among uranium and other hard rock miners have given relatively consistent risk estimates for the lung (UNSCEAR 1977). Uncertainties in radiobiological effectiveness of this radiation and the carcinogenic effect of smoking complicate the general applicability of these estimates. Definite evidence and correlation of bone cancer have been obtained from radium dial painters and from people who were administered radium for its supposedly therapeutic effect (Rowland et al 1971). Radiologists who were exposed to substantial occupational doses during the early part of this century, in spite of uncertainty of dose received, provide assessments
of excess over expected different fatal malignancies. From this relative frequency of leukemia as compared with all other malignancies, for more or less uniform irradiation of the whole body, can be obtained (Matanoski et al 1975).

5. Somatic risk estimation

5.1 Dose-response models

We have seen above that the cancer incidence data available is at high doses and high dose rates. The need, however, is to estimate the risk of cancer induction at very low doses and dose rates. Therefore, it becomes necessary to construct dose response models to fit the available data and extrapolate it to low doses. It is this modelling and extrapolation which has been a matter of controversy. In recent years, a general hypothesis for estimation of excess cancer risk in irradiated human populations, based on theoretical considerations, on extensive laboratory animal studies, and on limited epidemiological surveys has been developed which takes the complex quadratic form (NAS 1980):

\[ I(D) = (a_0 + a_1 D + a_2 D^2) \exp (-\beta_1 D - \beta_2 D^2), \]

where \( I \) is the cancer incidence in the exposed population at dose \( D \) (cGy).

This multicomponent dose-response curve contains initial upward curving linear \((a_1)\) and quadratic \((a_2)\) functions of dose, which represent the process of cancer-induction by radiation, and a modifying exponential function of dose which represents the competing effect of cell-killing at high doses \((\beta_1 \text{ and } \beta_2)\). The constant \( a_0 \) accounts for the natural incidence of cancer in the population. Attempts to fit the available data have demonstrated that for different radiation-induced cancers only certain of the parameter values of these constants can be theoretically determined. Therefore, simpler models for the dose-response relationship in the low-level dose range are advocated as shown in figure 6, namely,

(i) linear \( I(D) = a_0 + a_1 D \),
(ii) quadratic \( I(D) = a_0 + a_2 D^2 \),
(iii) linear quadratic \( I(D) = a_0 + a_1 D + a_2 D^2 \).

5.2 Risk projection models

The US National Academy of Sciences (1980), BEIR-III report and the United Nations Scientific Committee on Effects of Atomic Radiation (1977) report use two models: the absolute risk model and the relative risk model. The absolute risk is the expression of excess cancer risk due to radiation exposure as the arithmetic difference between the risk among those exposed and that occurring in the absence of exposure (figure 7). This model considers the latent period and the period of expression. The absolute lifetime risk co-efficient is expressed as the total number of excess cancer cases in the exposed population per unit collective dose. The relative risk is the expression of cancer risk due to exposure as the ratio of the risk among the exposed population to that occurring in the absence of exposures (figure 7). Therefore, the excess risk is a multiple of the spontaneous age-specific cancer rate in that cohort population. The greater the natural rate of cancer incidence in a population (e.g. ageing population) the greater will be the
Figure 6. Radiation risk models (source Goldman 1982).

Figure 7. Models of absolute and relative risk and age dependence. Xe is age at exposure and l is the minimum latent period (Source NAS 1980).
Table 4. Risk estimates for individual organs—cancer risk. Fatalities per rem per $10^6$ persons.

<table>
<thead>
<tr>
<th>Organ</th>
<th>UNSCEAR 1977</th>
<th>BEIR III 1980</th>
<th>ICRP 26 1977</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood (leukemia)</td>
<td>15–25</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>Breast</td>
<td>50</td>
<td>87</td>
<td>25</td>
</tr>
<tr>
<td>Lung</td>
<td>25–50</td>
<td>42</td>
<td>20</td>
</tr>
<tr>
<td>Thyroid</td>
<td>10</td>
<td>2.2 M</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.8 F</td>
<td></td>
</tr>
<tr>
<td>Bone</td>
<td>25</td>
<td>—</td>
<td>5</td>
</tr>
<tr>
<td>Stomach</td>
<td>10–15</td>
<td>15 M</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17 F</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>10–15</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Lower large intestine</td>
<td>10–15</td>
<td>—</td>
<td>10</td>
</tr>
<tr>
<td>Gonads</td>
<td></td>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>

susceptibility of the individuals comprising that population to radiation-induced cancer. Since epidemiological studies on human population are not complete as yet (not followed up to death) there can be wide differences in risk estimation by the two methods at any one period of follow-up. When the study is complete so that no more cancers occur in the studied population, both models should lead to the same numerical estimates of lifetime cancers though the risk distribution may be different in the exposed population. The BEIR-III report emphasised that some experimental and human data, as well as theoretical considerations, suggest that for exposure to low doses of low LET radiation such as x-rays and gamma rays, the linear model probably overestimates but can be used to define the upper limits of risk of most radiation-induced cancers in man. Similarly the pure quadratic model may be used to define the lower limits of risk from low dose, low-LET radiation. For exposures to high-LET radiation (neutron, alpha particles, etc) linear risk estimates are less likely to overestimate the risk and may, in fact, underestimate. The International Commission on Radiological Protection has used the linear extrapolation to estimate risks. Table 4 gives the risk estimates for individual organs by different bodies.

6. Genetic effects

The induction of genetic mutations by radiations has been known now for more than five decades. A large amount of information is available on the frequency with which certain hereditarily transmitted abnormalities are caused in the mouse, as well as in various vegetable and other animal forms, by irradiation even at low doses (UNSCEAR 1977; Oftedal and Searle 1980). In man, however no corresponding information has been obtainable on the rate of induction of harmful genetic effects by radiation. But the frequency of chromosomal aberrations is known in human blood lymphocytes cultured after exposure of the blood ‘in vitro’ or after irradiation received in vivo in the course of medical or other procedures at known dose. But these cannot form a basis for genetic risk estimation. The latter estimates must at present be based largely upon mice data of
radiation exposure required to double the normal frequency, the doubling dose. This
doubling dose for twelve different types of hereditary abnormality in male or female
mice have ranged from 0.4 to 2.6 with a mean of 1.4 (± 0.2, S.E) Gy (Searle 1977).

Genetic risk estimates have been obtained assuming that a comparable doubling
dose, taken as 1 Gy for low-LET low-dose rate radiation, is a characteristic of genetic
damage in mammals and may also apply to man (UNSCAR 1977). Knowing the normal
frequency of all genetic abnormalities that are maintained by mutation, a risk estimate
of major genetic harm expressed in the first generation has been derived as
0.9 x 10^-2 Gy^-1 with 0.35 x 10^-2 Gy^-1 in the second, and a total of 3.2 x 10^-2 Gy^-1
in all generations (Ofstedal and Searle 1980). UNSCER (1977) estimated 0.65
x 10^-2 Gy^-1 for first generation and a total of 1.85 x 10^-2 Gy^-1 for all generations.
ICRP (1977a) has used a value of 10^-2 Gy^-1 for the first two generations and 2
x 10^-2 Gy^-1 for all generations.

7. Risk limitation through system of dose limits in radiation protection

It is believed that like genetic mutations, for cancer induction also the initiating
processes are chromosomal, since the abnormal reproductive behaviour of transformed
cells must be transmitted in subsequent cell divisions (Pochin 1980; Rossi and Kellerer
1972). The dose-effect relationship is thus likely to have the same general form as for
genetic effects. Cancer induction and genetic effects are therefore classed as stochastic
effects; where there is no threshold and there is a probability of induction at all doses
however low. Non-stochastic effects on the other hand occur if their threshold is
exceeded and their severity depends on the amount of dose.

Non-stochastic effects are prevented by ensuring that no tissue receives annual doses
which reach threshold values after a lifetime. ICRP (1977a) has recommended that no
tissue shall receive more than 0.05 Sv annually under conditions of occupational
exposure and 0.5 Sv per year in members of the general public.

The stochastic effects are not preventable but their frequency of occurrence can be
limited. It is important, therefore, that unnecessary exposure shall be avoided and
unavoidable exposure shall be as low as reasonably achievable (ALARA). The system of
dose limits that is adopted, therefore, indicate maximum values which should never
ordinarily be exceeded. The radiation safety of an occupation depends on the risks
corresponding to the average annual exposure received. In order to compare the safety
of occupations involving radiation exposures with those not involving, accounts should
be ideally taken of the whole range of occupational injuries, diseases, fatalities, the
mutagenic or carcinogenic exposures if any, and a variety of transient or permanent
disabilities, etc (ICRP 1977b). Radiation exposures which amount to an average annual
value of 5 mSv may cause a component of harm equal to about 3–4 accidental deaths
annually per 100,000 workers. If on the other hand the working conditions are such that
the annual dose limit of 50 mSv was commonly approached each year by most workers
leading to an average dose of 30 mSv per year, then the radiation harm would be
equivalent to about 20 accidental deaths annually per 100,000 workers (Pochin 1980).

A coherent system of dose limits must have knowledge of and suitably equate risks of
irradiating individual body organs, relative to that of uniform whole body exposure to
ensure compliance with the limits. To achieve this the ICRP (1977a) has used weighting
factors based upon the risks of fatal cancers of different organs or genetic risks. Thus,
radiation protection has come a long way from the earlier empirical though safe, dose limits to the recent more scientifically derived sound basis. A lot more data and quantitative information on several aspects of radiation effects are required to make radiation protection still more authoritative.

Acknowledgements

The author thanks Shri S D Soman, Associate Director, Radiological Group, BARC for his keen interest in this work.

References

Baral E, Larsson L E and Mattsson B 1977 Cancer 40 2905
Cohen B L 1981 Health Phys. 40 305
Conrad R A et al 1975 A 20-year review of medical findings in a Marshallese population accidentally exposed to radioactive fallout, Rep. BNL 50424, Brookhaven National Lab., Upton, New York, USA
Darby S L and Keeney R L 1981 Risk Analysis 1 217
Fischhoff B, Slovic P and Lichtenstein S 1979 Environment 21 17
Glasstone S and Sesonske A 1963 Nuclear reactor engineering (New Jersey: Van Nostrand) (By permission: USAEC)
Goldman M 1982 Western J. Med. 137 6
ICRU 1980 Radiation quantities and units, ICRU Rep. No. 33, ICRU, Washington DC, USA
Jain V K and Soman S D 1978 SI units in radiation protection, BARC Rep. 964, BARC, Bombay
Lowe W E 1982 A-bomb survivor dosimetry update. UCRL-86275, Lawrence Livermore National Lab., Livermore, Calif., USA
NAS 1980 The effects on populations of exposure to low levels of ionizing radiations. Rep. of Committee on Biological Effects of Ionizing Radiations, National Academy of Sciences, BEIR III
Oftedal P and Searle A G 1980 J. Med. Genet 17 15
Pochin E E 1980b Atomic Energy Rev. 18 779
Rossi H H and Kellerer A M 1972 Science 175 200
Shore R E, Albert R E and Pasternack B S 1976 Arch. Environ. Health 31 21
Smith P G and Doll R 1976 Br. J. Radiol. 49 224
Starr C 1969 Science 165 1232
Starr C 1981 Risk Analysis 1 117
Stewart A and Kneale G W 1970 Lancet 1 1185
Electron energy loss spectroscopy (EELS) of organic molecules in vapour phase: Design and fabrication of an indigenous EEL spectrometer*

M S HEGDE, V JAYARAM, P VISHNU KAMATH and C N R RAO†
Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560 012, India

Abstract. An indigenous electron energy loss spectrometer has been designed and fabricated for the study of free molecules. The spectrometer enables the recording of low-resolution electronic spectra of molecules in the vapour phase with ready access to the vacuum ultraviolet region. Electron energy loss spectra of aliphatic alcohols and carbonyl compounds as well as of benzene derivatives have been recorded with the indigenous spectrometer and the electronic transitions in these molecules discussed.

Keywords. Electron energy loss spectroscopy; vacuum ultraviolet spectroscopy; EELS.

PACS No. 33·20; 34·80

1. Introduction

Electron energy loss spectroscopy (EELS) provides an effective means of investigating electron states of free atoms and molecules (Celotta and Huebner 1979; Rao et al 1980). In this technique, a primary electron beam of energy \( E_i \) collides with atoms or molecules and the inelastic collisions cause excitations within the sample. The kinetic energy \( E_f \) of the inelastically scattered electrons is measured and the difference, \( \Delta E = E_i - E_f \), directly gives the energy absorbed by the molecule during the collision process (figure 1). EELS is similar to optical absorption spectroscopy, but the advantage of EELS lies in the fact that it can be utilised to study optically forbidden transitions (Kuppermann et al 1979; Koerting et al 1984). The phenomenon of electron exchange (Oppenheimer 1928) whereby the scattered electron is different from the incident one, enables us to study spin-forbidden transitions. Symmetry-forbidden transitions in molecules can also be observed since the incident electron can polarize the electron cloud distribution of the molecule during collision.

What is specially significant about EELS is that it gives us ready access to the vacuum UV region of electronic spectroscopy and this in conjunction with photoelectron spectroscopy provides the complete molecular energy level diagram of the molecule under study. Since an EELS instrument is relatively simple and inexpensive to fabricate, it can be regarded as a laboratory vacuum UV spectrometer of low resolution for studies

---

* Contribution No. 274 from the Solid State and Structural Chemistry Unit, dedicated to Dr. Raja Ramanna on his 60th birthday.
† To whom all correspondence should be addressed.
involving identification and characterization of molecules in the vapour phase. EELS has the disadvantage of low resolution, a feature common to all techniques of electron spectroscopy. In this paper we report the design and fabrication of a simple EEL spectrometer devoted to the study of atoms and molecules in the gas or vapour phase. We discuss the EEL spectra of several organic molecules recorded with this indigenous spectrometer, having calibrated the spectrometer with gases such as He, O₂ and N₂.

2. Design of the spectrometer

The spectrometer consists of an electron beam source, a monochromator, electron lenses, sample chamber, analyser and a detector. As most of the electronic excitations in the uv and vacuum uv regions lie in the energy range of 4–10 eV, the energy of a typical source should be about three to five times this energy. We have therefore selected a source energy in the range of 30–70 eV. The design criteria for fabrication of the spectrometer were based on the three important factors discussed below.

2.1 Incident electron beam intensity

As the electron beam passes through each of the components of the spectrometer, there is a certain loss in intensity. This loss is minimized by the use of appropriate electron optics; the number of inelastically scattered electrons would also be small, being typically two orders of magnitude less than the incident beam. In view of these factors, it is necessary to have an intense elastic peak of at least $10^5$ counts/sec at the mouth of the detector. This would correspond to a current of $10^{-14}$ amps. Assuming a loss in intensity of the order of $10^4$ in the monochromator, $10^3$ in the sample chamber and a further loss of 10 in the analyser, the total loss over all the components would be of the order of $10^8$. The electron gun source must therefore be capable of emitting in the range of 1–10 microamps in the energy range 30–70 eV. The electron source was designed with these specifications in view.
2.2 Resolution

The desired resolution of a spectrometer of this kind meant for the study of organic molecules (operating under low resolution), is approximately 400 meV for a primary energy of about 50 eV. Resolution of this order can be obtained using spherical sector analysers (Kuyatt and Simpson 1967). The resolution $\Delta E/E$ of spherical sector analysers is given by the formula $\Delta E/E = S/2R_0$ where $S$ is the slit width and $R_0$ is the mean radius of the analyser. If we use a convenient slit width of 1 mm, an analyser of a mean diameter of 125 mm is needed.

2.3 Vacuum conditions

Since the spectrometer is devoted to the study of atoms or molecules in the gas or the vapour phase, it is clear that two separate and independent pumping systems are required. One provides the base vacuum required for monochromator and the spectrometer chambers, while the other provides differential pumping for the stream of sample molecules. The sample pressure required to record an EEL spectrum depends on the nature of the sample, as well as the length of the sample chamber. The sample pressure must be such that the mean free path of the electrons for electron impact excitation must be less than the length of the collision chamber. The mean free path $\lambda$ is defined as $\lambda = 1/na$, where $n$ is the number of molecules per c.c. and $\sigma$ is the inelastic electron impact collision cross-section. Assuming $\sigma$ to be in the range $10^{-15}-10^{-16}$ cm$^2$ (McDaniel 1964) for a $\lambda$ of approximately 25 mm, the sample pressure required is of the order of $10^{-4}-10^{-3}$ torr. The desired base vacuum of the spectrometer on the other hand is of the order of $10^{-6}$ torr. Hence the crucial need would be for an effective differential pumping of the sample. The vacuum system was designed keeping these requirements in view.

3. Fabrication of the spectrometer

3.1 Electron source

A schematic diagram of the electron gun used in the spectrometer is shown in figure 2. A thoriated tungsten filamet of 0.1 mm dia is mounted on a pair of Ni rods. The filament is

![Figure 2](image_url)
resistively heated with a constant current-stabilised power supply fabricated in the laboratory. The electron beam current is controlled by using a negatively biased grid. The gun is mounted on a 63 mm diameter brass flange and enclosed in a stainless steel casing as shown in figure 2. Electrons with energy of 30–70 eV were obtained in this manner.

When the filament is fitted to a reflecting electrode $R$, which is at the same voltage as the filament, an emission current of 5–10 $\mu$ amp can be obtained.

3.2 Monochromator and analyser

The monochromator as well as the analyser consist of two hemispherical sectors of mean diameter 125 mm. The hemispherical sectors have been machined out of Al blocks to an accuracy of $\pm 0.025$ mm. The hemispherical sectors are gold-coated by vacuum deposition employing planetary rotation, to ensure uniform coating of the spherical surfaces. The sectors are then mounted on a 6 mm thick aluminium support plate with suitable electrical insulation (figure 3); the electrical insulations are also shown in the figure.

![Figure 3. Schematic illustration of the analyser assembly. Shaded areas are machined out of teflon. IS: inner sphere; OS: outer sphere; SP: support plate; MF: main flange. Inset shows the details of insulation employed in the assembly of the analyser.](image)

The performance of the spherical sector analysers was measured over a wide range of electron beam energies (25–100 eV) of interest to us. The analyser response was found to be linear over the 30–70 eV range. Analysers of mean diameter 114 mm and 139 mm, have also been machined and tested for their linear response. The analyser performance was found to improve on coating the slits with gold.

3.3 Electrostatic lenses

An electrostatic lens systems was machined as described elsewhere (Spangenberg and Field 1943; Lassettre et al 1963). The lens assembly (figure 4) consists of a pair of
identical hollow cylindrical tubes mounted co-axially in an outer jacket and insulated electrically from one another, the tubes being separated from each other by a distance equal to one-tenth of their diameter. The lens system mounted on the “out” slit of the monochromator, acts as a collimator and helps to focus the beam emerging from the monochromator into the sample chamber. The cylindrical lenses were machined from stainless steel rods of suitable dimensions to an accuracy of ± 0.025 mm and coated with gold. The lens system was found to work with negligible chromatic aberration.

3.4 Sample chamber

The sample chamber was constructed out of a stainless steel bellow 25 mm in diameter and 25 mm in length. It was fitted between two brass collars, which were in turn welded to the sample input and output lines. The gaseous sample is introduced through a sensitive needle valve. When the sample pressure in the sample chamber was $10^{-3} - 10^{-4}$ torr, the base vacuum in the spectrometer chambers increases from $2 \times 10^{-6}$ torr to $4 \times 10^{-6}$ torr. The electron beam emerging from the lens system meets the stream of sample molecules at right angles in the sample chamber.

The flexibility of the bellow permits us to collect the scattered electron current along off normal angles. A study of the angular dependence of the scattered electron current can provide information on the shapes and symmetries of orbitals.

3.5 Detector and ratemeter

A Mullard channel electron multiplier (B419BL) was used to amplify the electrons coming from the energy analyser; this is the only imported component in the spectrometer. The electron multiplier responds to an input of one electron upto about $10^8$ electrons and its duration (full-width at half-maximum) is about 10 nanosecond. The maximum operating voltage of the multiplier is around 3.5 kV, the nominal gain at 2.5 kV being about $1.1 \times 10^8$ at a nominal resistance of $5 \times 10^9$ ohm.

The high impedance transducer (detector) such as the electron multiplier requires an amplifier which converts a transfer of charge into a change of voltage. In the ratemeter (figure 5), we used a hybrid charge-sensitive preamplifier-discriminator (AMPTEK A-101 PAD) to convert pulses of charge to pulses of voltage. This is capable of handling counts upto a maximum of $4 \times 10^6$ with pulse pair resolution of 250 nsec. The input threshold could be varied from $1.6 \times 10^{-13}$ coulomb to $15 \times 10^{-12}$ coulomb.

The output from the charge-sensitive amplifier is fed to the decade counter and monostable multivibrator. This positive pulse is fed to the voltage level translator.
which produces the constant 12 V positive pulse. This constant height pulse is fed to the 'frequency to voltage' converter and buffer. Reference voltage is used to bias the voltage level translator and also to introduce zero bias to 'frequency-to-voltage' converter. The output dc voltage from the ratemeter is fed to the Y axis of the X-Y recorder which directly gives the number of counts per second.

3.6 General assembly and the vacuum system

The spectrometer consists of two main chambers 241 mm in diameter and 228 mm in length fitted to brass collars of 304 mm diameter. The main chambers are made by rolling 3 mm stainless steel plates into cylinders with all the joints being accomplished by silver brazing. Both the chambers have been provided by two more sheets of μ-metal for shielding the earth's magnetic field.

The general assembly of the spectrometer is shown in figure 6 and the electronic control in figure 7. The source, monochromator and the electrostatic lens are fitted to a 300 mm diameter Al flange 12 mm in thickness, and is fitted to the first chamber. The monochromator chamber is mounted on a 150 mm diameter oil diffusion pump which is backed by a 300 lit/min rotary pump. This consists of the main pumping system and is capable of providing an ultimate vacuum of better than $2 \times 10^{-6}$ torr, with a total pump down time of less than 30 min.

The analyser and detector are mounted on a 300 mm diameter Al flange of 12 mm thickness. This flange is fitted to the analyser vacuum chamber. The analyser chamber is also connected to the main vacuum system through a 25 mm diameter flexible tubing. The collision chamber is fitted between the two spectrometer chambers linking the monochromator output to the analyser input. The collision chamber is fitted to a 100 mm oil diffusion pump backed by a 300 lit/min rotary pump for differentially pumping the sample.

Both the spectrometer chambers are provided with pirani and hot cathode ionisation gauges for measurement of vacuum. The sample pumping line is also fitted with separate gauge heads for measuring the sample pressure. A photograph of the instrument in final assembly is shown in figure 8.
4. Operation of the spectrometer

The vacuum system is first switched on. On achieving the base vacuum of the system ($2 \times 10^{-6}$ torr), the source gun is put on and degassed for 10 to 15 min. The
collision chamber is pumped through the sample pumping line till the base vacuum in the sample pumping line ($4 \times 10^{-6}$ torr) is of the same order as that in the main pumping system. The source energy is selected to be in the 50–70 eV range and the monochromator and lens voltages are suitably altered until the Faraday collector C (see figure 6) mounted at the mouth of the analyser collects a maximum beam current ($\sim 10^{-12}–10^{-9}$ amps). This beam enters the analyser. A high voltage of 2.2–2.5 kV is applied to the multiplier and the ratemeter is switched on. The analyser voltage is scanned using a DC ramp. The spectra are recorded by scanning the voltage across the analyser by keeping the retard voltage at zero. We can also scan the retard voltage applied to the analyser mounting plate by applying a constant DC voltage to the hemispheres. The elastic peak is first recorded at low sensitivity of the ratemeter and having found the elastic peak, the sample is introduced into the collision chamber, until the sample pressure rises to $10^{-3}$ torr. The elastic peak height drops to half its original value at this juncture and in order to detect the loss peaks, the ratemeter is then fixed to high sensitivity (typically 300 to $10^3$ times higher than the sensitivity used to record the elastic peak).

5. Performance of the spectrometer

The spectrometer is readily calibrated by recording the spectra of standard gases such as He, N$_2$, O$_2$ and benzene vapour. The spectrum of He is shown in figure 9a, where peaks due to excitations into the various excited states are clearly assigned. The peak due to the $1^1S \rightarrow 2^1S$ excitation can also be assigned, thus demonstrating the versatility
of the spectrometer in detecting optically forbidden transitions. The optically allowed transitions to the $2^1P$, $3^1P$ and $4^1P$ excited states are seen as prominent peaks and are eminently suited for calibration. In figure 9b, we show the spectra of pure N$_2$, pure O$_2$ and air. The spectrum of air appears as due to a mixture of N$_2$ and O$_2$. The electronic excitations responsible for these peaks are described in the literature (Lassettre et al 1963; Kuppermann et al 1968).

At a primary beam energy of 50 eV, the fwhm of the elastic peak is 360 meV. The resolution of the spectrometer is 0.72% at 50 eV and the uncertainty in measurement is ±0.05 eV.

Figure 9. (a) EEL spectrum of He gas recorded in the spectrometer. Inset shows the spectrum on a expanded scale, together with the assignment of the transitions.
6. Study of EEL spectra of organic molecules

6.1 Aliphatic alcohols and diethylether

In figure 10 we show the EEL spectra of a few aliphatic alcohols and diethylether. The absorption maxima observed in these compounds are listed in table 1. These molecules can show transitions due to the $n \rightarrow \sigma^*$ and $\sigma \rightarrow \sigma^*$ excitations (Rao 1975). The $n \rightarrow \sigma^*$ transition would be associated with lower energies and intensities compared to the $\sigma \rightarrow \sigma^*$ transitions which occur in the vacuum UV region with high extinction coefficients. Accordingly in EEL spectra, the first band occurs at 177 nm (7 eV) in CH$_3$OH which compares well with the UV absorption band of 174 nm reported by Tsubomura et al (1964). This band is due to the $n \rightarrow \sigma^*$ transition involving the oxygen lone pair. The first band in CH$_3$OH is quite intense and sharp, but as the alkyl group becomes bulkier, the first band becomes broader and less intense. The broadening of
the $n \rightarrow \sigma^*$ band is expected as bulkier alkyl groups provide a wide spectrum of excited $\sigma^*$ states. Secondary and tertiary alcohols exhibit the $n \rightarrow \sigma^*$ band at lower energies compared to the primary alcohols due to enhanced inductive effect of the alkyl groups in the former.

In addition to the $n \rightarrow \sigma^*$ band, we see additional bands at higher energies in EEL spectra of all the alcohols. These bands are broad and increase in complexity as the alkyl groups become bulky. These bands may be divided into two categories. The first category appears in the 158–151 nm (7.8–8.2 eV) range and is absent in CH$_3$OH. It appears as a distinct peak in C$_2$H$_5$OH and i-C$_3$H$_7$OH, but becomes a shoulder in the
Table 1. Absorption maxima in the vacuum ultraviolet region for some aliphatic alcohols and halocarbons (a).

<table>
<thead>
<tr>
<th>Compound</th>
<th>n → σ*</th>
<th>σ → σ*(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH₃OH</td>
<td>177(6-98)</td>
<td>141(8-78)</td>
</tr>
<tr>
<td>EtOH</td>
<td>190(6-49)</td>
<td>132(9-39)</td>
</tr>
<tr>
<td>C₂H₅OC₂H₅</td>
<td>187(6-62)</td>
<td>139(8-9)</td>
</tr>
<tr>
<td>isopr-OH</td>
<td>191(6-5)</td>
<td>139(8-9)</td>
</tr>
<tr>
<td>npr-OH</td>
<td>177(6-98)</td>
<td>156(7-94)</td>
</tr>
<tr>
<td>t-Bu-OH</td>
<td>184(6-74)</td>
<td>151(8-18)</td>
</tr>
<tr>
<td>n-BuOH</td>
<td>198(6-26)</td>
<td>158(7-82)</td>
</tr>
<tr>
<td>CH₂Cl₂</td>
<td>171(7-22)</td>
<td>151(8-18)</td>
</tr>
<tr>
<td>CHCl₃</td>
<td>177(6-98)</td>
<td>145(8-54)</td>
</tr>
<tr>
<td>CCl₄</td>
<td>174(7-09)</td>
<td>141(8-78)</td>
</tr>
</tbody>
</table>

(a) Values in parentheses are transition energies in eV.
(b) At low wavelengths, some Rydberg transitions may be mixed with the σ → σ* transitions.

higher members of this series. This feature is likely to be due to the σ → σ* transition; since it is absent in CH₃OH, we may assign it to the σ → σ* transition involving the C–C bond.

The second category of σ → σ* transitions occurs at wavelengths below 150 nm (> 8-2 eV). There is a band appearing in the 140–130 nm (8-8–9-5 eV) range and another at or below 130 nm; the first is probably due to the σ → σ* transition involving the C–H bond. Below 130 nm (> 9-5 eV) excitation into the Rydberg states probably occurs.

In diethylether, the n → σ* band of ethanol splits into two bands, one at lower and the other at higher energy relative to ethanol.

6.2 Carbonyl compounds

In figure 11 we show the EEL spectra of a few carbonyl compounds. The carbonyl group exhibits n → π*, n → σ* and π → π* transitions in the electronic spectra (Rao 1975); the n → π* transition has a low extinction coefficient and is not seen in the EEL spectra recorded by us. The first band seen in EEL spectra appears in the energy range 180–190 nm (6-5–6-8 eV) in acetaldehyde and acetone and this can be readily assigned to the n → σ* transition of the carbonyl group. In acetone, it appears as a sharp peak, while in acetaldehyde it is a shoulder on a higher energy peak. In methylacetate also it appears as a weak shoulder. This band is however not seen in formaldehyde, which shows a single sharp peak at 172 nm (7-2 eV) due to the π → π* transition. The π → π* transition occurs with high intensity in the 166–172 nm (7-2–7-5 eV) range in all the carbonyl compounds studied by us (see table 2). Below 155 nm (8-0 eV), carbonyl compounds show a broad band with a multiplicity of peaks due to the σ → σ* transitions.

In formic and acetic acids, the peak due to the π → π* transition occurs at 153 nm (Rao 1975), while in acetic anhydride it is around 166 nm (7-5 eV) just as in other carbonyl compounds. Formic acid shows no other peak, but both acetic acid and its anhydride exhibit a broad band around 125 nm (9-9 eV) due to the σ → σ* transitions of the methyl group (figure 12). Presence of several Rydberg states causes broadening
of the $\sigma \to \sigma^*$ band in the higher energy region. Both the acids show a slight broadening of the $\pi \to \pi^*$ peaks possibly due to dimerization.

**EEL** spectra of a series of aliphatic amides are shown in figure 13. The band due to the $\pi \to \pi^*$ transition shifts considerably from 168 nm (7.4 eV) in formamide to 202 nm
Figure 12. EEL spectra of formic and acetic acids together with the spectrum of acetic anhydride.

Table 2. Absorption maxima seen in the vacuum UV region for a series of carbonyl compounds\(^{(a)}\).

<table>
<thead>
<tr>
<th>Compound</th>
<th>(n \rightarrow \sigma^*)</th>
<th>(\pi \rightarrow \pi^*)</th>
<th>(\sigma \rightarrow \sigma^*)(^{(b)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCHO</td>
<td>—</td>
<td>172(7.22)</td>
<td></td>
</tr>
<tr>
<td>CH(_3)CHO</td>
<td>180(6.85)</td>
<td>169(7.34)</td>
<td>147(8.42)</td>
</tr>
<tr>
<td>CH(_2)COCH(_3)</td>
<td>192(6.44)</td>
<td>—</td>
<td>152(8.12)</td>
</tr>
<tr>
<td>CH(_2)COOCH(_3)</td>
<td>180(6.85)</td>
<td>166(7.46)</td>
<td>130(9.51)</td>
</tr>
<tr>
<td>(CH(_3)CO(_2))(_2)</td>
<td>—</td>
<td>166(7.46)</td>
<td>128(9.62)</td>
</tr>
<tr>
<td>CH(_3)COOH</td>
<td>—</td>
<td>153(8.1)</td>
<td>124(9.98)</td>
</tr>
<tr>
<td>HCOOH</td>
<td>—</td>
<td>166(7.45); 153(8.1)</td>
<td></td>
</tr>
<tr>
<td>HCONH(_2)</td>
<td>—</td>
<td>168(7.34)</td>
<td></td>
</tr>
<tr>
<td>HCONH(Me)</td>
<td>—</td>
<td>177(6.98)</td>
<td></td>
</tr>
<tr>
<td>HCON(Me)(_2)</td>
<td>—</td>
<td>202(6.13)</td>
<td></td>
</tr>
<tr>
<td>CH(_3)CON(Me)(_2)</td>
<td>—</td>
<td>194(6.37)</td>
<td></td>
</tr>
</tbody>
</table>

\(^{(a)}\) Values in parentheses are transition energies in eV.

\(^{(b)}\) Rydberg transitions may be mixed up with the \(\sigma \rightarrow \sigma^*\) transitions at low wavelengths.
(6.1 eV) in N-N dimethylformamide. Similar shifts are seen in the acetamide series as well (table 2).

6.3 Benzene derivatives

In figures 14 and 15, we show the EEL spectra of benzene and several of its mono- and disubstituted derivatives. Benzene exhibits three bands in the electronic absorption spectra, two of these in the ultraviolet region (203 and 260 nm) and one in the far
Figure 14. EEL spectra of a few benzene derivatives.

Table 3. Absorption maxima in vacuum UV region for some aromatic molecules\textsuperscript{(a)}.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$(nm)</th>
<th>$p$(nm)</th>
<th>$\alpha$(nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C$_6$H$_6$</td>
<td>184(6:74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toluene</td>
<td>181(6:86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-xylene</td>
<td>190(6:49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorobenzene</td>
<td>190(6:49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bromobenzene</td>
<td>184(6:74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anisole</td>
<td>189(6:55)</td>
<td>214(5:77)</td>
<td>271(4:57)</td>
</tr>
<tr>
<td>Benzaldehyde</td>
<td>198(6:25)</td>
<td>234(5:29)</td>
<td>271(4:57)</td>
</tr>
<tr>
<td>Aniline</td>
<td>190(6:49)</td>
<td>224(5:53)</td>
<td>278(4:45)</td>
</tr>
<tr>
<td>N-Methyl Aniline</td>
<td>196(6:32)</td>
<td>231(5:35)</td>
<td>286(4:33)</td>
</tr>
<tr>
<td>N-N-Dimethylaniline</td>
<td>202(6:13)</td>
<td>245(5:05)</td>
<td>278(4:45)</td>
</tr>
<tr>
<td>N-N-Diethylaniline</td>
<td>198(6:25)</td>
<td>257(4:81)</td>
<td>294(4:21)</td>
</tr>
<tr>
<td>p-Chlorotoluene</td>
<td>190(6:49)</td>
<td>224(5:53)</td>
<td></td>
</tr>
<tr>
<td>p-Chlorophenol</td>
<td>190(6:49)</td>
<td>221(5:59)</td>
<td>282(4:39)</td>
</tr>
</tbody>
</table>

\textsuperscript{(a)} $\beta$, $p$ and $\alpha$ bands correspond to the 183, 203 and 260 nm bands of benzene respectively. Values in parantheses are transition energies in eV.
The ultraviolet region (180 nm). The last of these is the most intense with an extinction coefficient of $10^5$ (Rao 1975). The high symmetry of the benzene molecule causes the lower energy bands to have low intensities. Consequently in EEL spectra recorded by us, we only see the 180 nm (6-9 eV) band as a sharp feature. In the mono- and di-substituted derivatives, the $D_{6h}$ symmetry of the benzene molecule is destroyed and we begin to see
the lower energy bands as well. This is exemplified in benzaldehyde, where the 234 nm (5.3 eV) band has comparable intensity with the 180 nm band. Substituents cause a red shift of the 180 nm band of benzene (table 3). The shift is specially significant in aniline and its methylated derivatives, where the amino group can donate electrons to the aromatic ring.

Acknowledgement

The authors are thankful to the Department of Science and Technology, Government of India and University Grants Commission, for support of the instrument fabrication and research reported in this paper.

References

Kuppermann A, Flicker W M and Mosher O A 1979 Chem. Rev. 79 77
Kuppermann A, Rice J K and Trajmar S 1968 J. Phys. Chem. 72 3894
McDaniel E W 1964 Collision phenomena in ionised gases (New York: John Wiley)
Oppenheimer J R 1928 Phys. Rev. 32 361
Rao C N R 1975 Ultraviolet and visible spectroscopy—Chemical applications, 3rd edn (London: Butterworths)
Spangenberg K R and Field L M 1943 Elec. Commun. 21 14
Neutron diffraction studies of transition metal nitrides

N S SATYA MURTHY†, S K PARANJPE, MARSONGKOHADI*, KURNIADI SUMAAMIDJAJA** and M R L N MURTHY
Nuclear Physics Division, Bhabha Atomic Research Centre, Bombay 400085, India
*Institute Teknologi Bandung, Indonesia.
**Pusat Penelitian Teknik Nuklir (Baden-Tenaga Atom Nasional) Bandung, Indonesia.

Abstract. The magnetic structures of Fe₄N and Mn₄N have been redetermined using neutron diffraction. The magnetic form factors, obtained from polarised neutron data have been shown to be different for the face-centred and corner atoms. A qualitative explanation of the structures of Fe₄N and Mn₄N has been provided from the shapes of the magnetic form factors.

Keywords. Neutron diffraction; magnetic form factor; nitrides

PACS No. 75-50

1. Introduction

Transition metal nitrides Fe₄N and Mn₄N exhibit interesting magnetic properties with ferro and ferrimagnetic ordering respectively. The crystal structure has space group Pm3m in which the metal atoms form a fcc matrix and the nitrogen atom occupies the interstitial body centre position (figure 1). The corner atom (M₁) has twelve metal atoms as its nearest neighbours at 2.69 Å whereas the face centre atom (M₁I) has two

Figure 1. Unit cell of transition metal nitride M₄N.

†Since deceased.
nearest neighbour nitrogen atoms at 1.9 Å. Because of this difference in their environments, the corner and the face centre atoms show different magnetic properties.

These compounds have been extensively studied for their electronic and magnetic properties. Following the pioneering work of Guillaud (1953) and Wiener and Berger (1955) on these compounds, Frazer (1958) undertook a neutron diffraction study of Fe₄N and concluded that it is ferromagnetic with magnetic moments of 2.73 µ_B and 1.77 µ_B for the corner and the face centre atoms respectively. The Mössbauer measurements (Shirane et al 1962) show that the hyperfine fields at the corner and face centre atoms are approximately proportional to their magnetic moments of 3 µ_B and 2 µ_B respectively and the isomer shifts are in line with 3d⁷ 4s and 3d⁸ 4s configurations for the two sites. These configurations are derived from the assumption that the nitrogen acts as an electron donor to the face centre iron atoms. However, the electron diffraction (Nagakura 1968) yields the electronic configuration as Fe⁰(Fe^{1/3})₃N.

Neutron diffraction studies on polycrystalline Mn₄N (Takei et al 1960; Takei et al 1962) show that it has a ferrimagnetic ordering with corner atom moment of 3.85 µ_B aligned antiparallel to the three face centre moments of 0.9 µ_B at 77°K. On the basis of neutron diffraction and neutron polarisation analysis (Bouchard 1968; Fruchart et al 1979) it has been claimed that Mn₄N has a noncollinear component of spin.

Mn₄N has been subjected to a variety of experimental measurements (Hisashi et al 1967; Mamoru 1966; Fruchart et al 1979) and it has exhibited anomalous magnetic and electronic properties. These observations have led to the band model calculations (Labbe and Jardin 1982; Jardin and Labbe 1975, 1983) based on a strong crystal field at the Mn site and a resonance between Mn and N electronic levels.

Regarding the magnetic structure of Fe₄N and Mn₄N, the existing literature appears to show an incomplete appreciation of the importance of the difference in the environments of the two species of metal atoms in these structures. The fact that the two atoms have significantly different environments leads to a difference in their magnetic moments. It should immediately follow that in the neutron experiments the atoms must also have different magnetic form factors. All the analysis so far have used the same magnetic form factors for the two sites, which leads to a serious flaw in the conclusions reached.

The present measurements are aimed at understanding not only the magnetic structures but the magnetic form factors in these compounds. The results on magnetic structures agree with the earlier results (Frazer 1958; Takei et al 1960). Interestingly, the magnetic form factors for the face centre Fe and Mn atoms in Fe₄N and Mn₄N respectively are found to be significantly different from those of the corner atoms.

2. Experimental

2.1 Sample preparation

Fe₄N was prepared by passing dry ammonia over 4N pure iron sponge, kept at 475°C, for 3 hr. The mixture of Fe₄N and Fe₃N thus formed, was heated in vacuum at 450°C for 4 hr. The resulting product was cubic γ' phase of Fe₄N, which was confirmed by x-ray analysis. For Mn₄N, nitrogen gas free of oxygen was passed over 4N pure manganese powder kept at 600°C. It was found necessary to crush the resulting product and repeat the heat treatment a couple of times to ensure a total reaction. X-ray analysis confirmed the formation of single phase Mn₄N.
2.2 Neutron diffraction measurements

The unpolarised neutron diffraction measurements were conducted at the TRIGA MARK II reactor PPTN Bandung, with neutron wavelength of 1.07 Å. Polarised neutron diffraction patterns were recorded using the polarised neutron spectrometer at the CIRUS reactor, BARC. This instrument has a wavelength of 0.92 Å and the polarisation and flipping efficiencies of 97.5% and 99.9% respectively.

3. Results and discussion

3.1 \( \text{Mn}_4\text{N} \)

The analysis of unpolarised neutron diffraction profiles shows that the structure is ferrimagnetic with magnetic moments of 3.50 \( \mu_B \) and 0.89 \( \mu_B \) for the corner and the face centre atoms respectively. The polarised neutron diffraction patterns are given in figure 2. Unpolarised neutron data is also shown for comparison. The depolarisation of the neutron beam in the sample was obtained by inserting the sample before the analyser crystal and measuring the polarisation of the neutron beam. From the polarised neutron intensities, the magnetic form factor for the face centred Mn atom was determined using Mn\(^+\) form factor (Watson and Freeman 1961) for the corner atom. Figure 3 shows the experimental form factor for Mn\(_{\text{II}}\) along with the Mn\(^+\) and Mn\(^{++}\) (Watson and Freeman 1961) form factors. The form factor for the Mn\(_{\text{II}}\) is sharper than that of Mn\(_{\text{I}}\). This suggests that the unpaired electron distribution of Mn\(_{\text{II}}\) is much more spread out than that of Mn\(_{\text{I}}\).

This is in agreement with the band model (Jardin and Labbe 1975, 1983; Labbe and Jardin 1982) for the transition metal nitrides and carbides. The model considers the

![Figure 2. Polarised and unpolarised neutron diffraction patterns of Mn\(_4\)N.](image-url)
Figure 3. Magnetic form factor of face centre Mn atom in Mn₄N.

Figure 4. Crystal field levels of face-centred Mn (d-orbitals) and N (p-orbitals). $E_F$ is Fermi energy (Labbe and Jardin 1982).
crystal field effects at Mn\textsubscript{II} site. The site symmetry of Mn\textsubscript{II} is $D_{4h}$ with two nitrogens as nearest neighbours. The 3d-levels of Mn\textsubscript{II}, in this strong crystal field, split into $a_{1g}$, $e_g$, $b_{1g}$, and $b_{2g}$ sublevels as shown in figure 4. Due to the large crystal field splitting, Hund's rule is not strictly valid and the magnetic moment is reduced. Further the $p$-level of nitrogen is in resonance with the $d_{e_g}$ level of Mn\textsubscript{II} resulting in the expansion of the unpaired electron density.

The experimental observations are thus in qualitative agreement with the theoretical model. The expanded moment density of Mn\textsubscript{II} also implies an overlap of the unpaired electron distributions of Mn\textsubscript{I} and Mn\textsubscript{II}, favouring a ferrimagnetic alignment.

3.2 Fe\textsubscript{4}N

The unpolarised neutron diffraction pattern of Fe\textsubscript{4}N was recorded using neutrons of wavelength 1.07 Å. The intensities of various reflections were measured by applying a magnetic field parallel to their respective scattering vectors. The magnetic structure, obtained using this data, gave a ferromagnetic alignment with magnetic moments of 2.73 $\mu_B$ and 1.77 $\mu_B$ for Fe\textsubscript{I} and Fe\textsubscript{II} atoms. This is in agreement with the earlier neutron results on Fe\textsubscript{4}N (Frazer 1958).

![Figure 5. Magnetic form factor of face-centred Fe atom in Fe\textsubscript{4}N. Unpolarised neutron data (open circles) gives an indication of the results discussed in the text.](image)
The polarised neutron diffraction data on Fe$_4$N was analysed on lines similar to those of Mn$_4$N. The magnetic form factor for the Fe$_{ii}$ atom has been obtained from the intensities using Fe$^+$ (Watson and Freeman 1961) form factor for Fe$_i$ and is shown in figure 5. The calculated Fe$^+$ and Fe$^{++}$ form factors (Watson and Freeman 1961) are also shown for comparison.

The form factor for Fe$_{ii}$ is significantly expanded as compared to that of the corner Fe$_i$ atom implying a contracted moment density for the Fe$_{ii}$ atoms. The unpaired electron densities of Fe$_i$ and Fe$_{ii}$ are thus well localised and do not overlap leading to a ferromagnetic ordering between Fe$_i$ and Fe$_{ii}$. The reduction in the moment is probably due to the strong crystal field effects at the Fe$_{ii}$ sites modifying the electron density contributing to the magnetic moment.

4. Conclusion

The polarised neutron diffraction data analysed on the basis of different form factors for the two nonequivalent sites thus give a significant difference in the unpaired electron densities at the corner and the face-centred atoms. The experimentally determined moment densities have qualitatively been explained on the basis of an existing theoretical model.

References

Bouchard J P 1968 Ann. Chim. Fr. 3 82
Frazer B C 1958 Phys. Rev. 112 751
Guillaud C P 1953 Rev. Mod. Phys. 25 119
Wiener G W and Berger J A 1955 J. Metals 7 360
Melting in two dimensions—the current status

G VENKATARAMAN* and D SAHOO
Reactor Research Centre, Kalpakkam 603 102, India

Abstract. The current status of the controversy relating to melting in two dimensions is surveyed. To begin with, a review is given of the seminal work of Kosterlitz and Thouless. This is followed by a discussion of the modifications introduced by Nelson and Halperin. The search for the continuous transitions and the intermediate hexatic phase predicted by these theories is then described, covering both the laboratory as well as simulation experiments. Alternate viewpoints to the $kT$ theory aired recently in the literature are also briefly examined. The paper concludes with an outlook for the future.

Keywords. Two-dimensional melting; computer simulation; Kosterlitz-Thouless theory; grain-boundary mechanism.

PACS No. 64-60

1. Introduction

As is well known, melting is a first-order transition. This of course refers to three-dimensional solids, but then we do live in a three-dimensional world. Be that as it may, it is interesting to speculate on the nature of melting of lower dimensional solids. The question is not purely academic since many solids actually behave as if they were of lower dimensionality.

Our attention in this paper will be directed to the melting of 2D solids, a topic which during the past decade has been the subject of active research as well as of lively controversy (as yet unsettled). After discussing why 2D melting is of interest per se, we will explain what the controversy is all about. We will then make a broad survey of the various attempts to shed light on the controversy, and wind up with an outlook for the future. Parenthetically we remark that although our focus is restricted to 2D melting, this topic actually forms a part of the general area of phase transitions in 2-dimensions, a subject with considerable implications far beyond condensed matter physics.

2. Some preliminaries

It is pertinent as well as useful to commence our discussion with a few general remarks on order in lower-dimensional systems. Nearly fifty years ago, Peierls (1935) and Landau (1937) demonstrated that long-range order is impossible in a 2D solid in the sense that $\langle (r - \bar{r})^2 \rangle$, the mean-square fluctuation of the distance $r$ between two atoms separated by average distance $\bar{r}$, increases logarithmically with the size of the system, in contrast to the finite value one obtains for a three-dimensional solid. In other words, crystalline order is rendered impossible due to fluctuations. One can understand this

Jawaharlal Nehru Fellow
intuitively by considering the number of nearest neighbours in a simple cubic, a square and a one-dimensional lattice. Whereas in the three-dimensional lattice there are 6 nearest neighbours to "hold back" every atom, there is less "check" in lower dimensional systems, facilitating the disruption of order. Slightly more formally, if \( \langle u^2(R) \rangle \) denotes the mean square fluctuation of atom at \( R \), then from a consideration of the thermal agitation caused by phonons it is found (Peierls 1935) that

\[
\langle u^2 \rangle \sim \int dq/q^2.
\]

This integral converges in three dimensions but diverges in lower dimensions, implying the absence of long-range order. In physical terms, long wavelength phonons prevent the solid from fully attaining a crystalline structure.

While the above line of reasoning is no doubt persuasive, one might legitimately wonder why it is that the famous two-dimensional Ising model does not fall into this slot but shows long-range ordering instead. Indeed, stimulated by such a line of thought, there were several studies in the sixties concerning the nature of order in various 2D systems. Mermin and Wagner (1966) for example, proved that there is no spontaneous magnetization in a 2D magnet with spins having more than one degree of freedom (for comparison, the Ising spin has only one degree of freedom). Shortly thereafter, Hohenberg (1967) demonstrated the absence of long-range order in a 2D Bose fluid.

Even as evidence began to accumulate concerning the absence of long-range order (LRO) in several 2D systems, Stanley and Kaplan (1966) discovered by analyzing the high-temperature series expansion that the susceptibility of the 2D planar magnet (the so-called XY model) diverged at a particular temperature, suggestive of a phase transition. This was quite intriguing for, till then, one was used to a phase transition accompanied by the onset of order. Did the work of Stanley and Kaplan imply that phase transitions were possible without requiring LRO? It was at this juncture that Kosterlitz and Thouless (1973) put forth their famous idea of phase transitions accompanied by a change of "topological order" rather than the conventional LRO. Understandably, great excitement was aroused, and various 2D systems were experimentally studied to check out these novel concepts. While in 2D superfluid a Kosterlitz-Thouless (KT) transition was observed as predicted (Bishop and Reppy 1978), the results on melting have been quite controversial and it is these that we shall be dwelling upon.

### 3. The Kosterlitz-Thouless theory

#### 3.1 Aspects of the 2D XY model

As a prelude to the KT theory, we shall briefly consider the 2D XY model in order to highlight the essential physical content of the KT mechanism. We begin by labelling systems using two integers \((d, n)\), denoting respectively the dimensionality of the system and that of the order parameter (i.e. the number of independent components in it). In this scheme, the Ising model would be a \((2, 1)\) system, the Heisenberg model investigated by Mermin and Wagner (1966) would be a \((2, 3)\) system and so on. The models of current interest (including the XY model) belong to the \((2, 2)\) family.
The 2D $XY$ model refers to a system of spins constrained to rotate in a plane. The spins are often taken to be organized on a lattice. However, one eventually goes to a continuum limit, whereupon the lattice structure can be ignored. Observe that the spin vector can continuously rotate and has an infinity of orientations to choose from, in contrast to the Ising spin which has only two options. This leads to an important difference in the group structure of the Hamiltonian. Whereas the Ising Hamiltonian has the discrete symmetry $Z_2$ (group of two elements), the Hamiltonians of some of the systems we shall be discussing all have the continuous symmetry $O(2)$. Particle physicists have strong interest in 2D $O(n)$ models of various types (Barber 1980).

Passing reference must also be made to the Potts model (1952) which is a generalization of the Ising model wherein the order parameter can take $q$ distinct values. For $q = 2$ (in 2D), the model reduces to the Ising model while in the limit $q \to \infty$ one recovers the $XY$ model. The Potts model thus interpolates between the $Z_2$ and $O(2)$ models.

Returning to the $XY$ model, its Hamiltonian may be expressed as

$$H = -J \sum_{\langle ij \rangle} S_i \cdot S_j,$$

where $J > 0$ and the sum $\langle ij \rangle$ is over-nearest neighbour pairs only (on a square lattice, say). Introducing for convenience the complex notation

$$\psi = S_x - i S_y = |S| \exp(i\theta),$$

and taking $|S| = 1$,

$$H = -J \sum_{\langle ij \rangle} \cos(\theta_i - \theta_j).$$

At low temperatures where spins on neighbouring sites are expected to be highly correlated, we may expand the cosine and retain only the quadratic term, obtaining

$$H = \frac{J}{2} \sum_{\langle ij \rangle} (\theta_i - \theta_j)^2.$$

The above Hamiltonian corresponds to the spin wave approximation (Wegner 1967) and is the starting point for the calculation of various physical properties. Of special interest is

$$C(r) \equiv \langle \psi(0) \psi^*(r) \rangle,$$

the correlation function. Analysis shows

$$C(r) = r^{-\eta(T)} \quad \text{for} \quad r \to \infty,$$

where

$$\eta(T) = k_B T / 2\pi J.$$
would terminate at a finite temperature $T_c$ (see figure 1c). Below that temperature there is power law behaviour while above there is exponential decay. The transition is thus characterized by changes in the correlation function rather than by the appearance or disappearance of LRO. As we shall presently see, these changes are linked to the topological order in the system.

### 3.2 Role of topological defects

How do topological defects alter the character of the system at $T_c$ as sketched in figure 1c? To answer this question, consider the vortex in figure 2a. This and its companion in figure 2b are outside the purview of the spin wave approximation. Observe that either of these vortices is capable of strongly disturbing spins far from the region of the singularity. By contrast, in the spin wave approximation there are no such long-range perturbing agencies (see figure 2c). However, with or without vortices, $C(r)$ always decays as $r \to \infty$ i.e. there is no LRO.

Vortices are like charges, and can be labelled with integers $N$ positive and negative. Those in figure 2 are the simplest, with $|N| = 1$. The energy of an isolated vortex may be calculated from (2) by using the continuum approximation, and is found to be (Kosterlitz and Thouless 1973)

$$E_{\text{vortex}} = E_c + \int_{a_0}^R 2\pi r dr \frac{J}{2} |\nabla \theta|^2$$

$$= E_c + \pi J \ln \left( \frac{R}{a_0} \right).$$

Here $R$ is the radius of the system and $a_0$ is a distance of atomic dimensions characterizing the core. $E_c$ is the core energy associated with the region $r < a_0$. Notice that the energy of an isolated vortex tends to infinity if the radius of the system goes to
Melting in two dimensions

Figure 2. Some possible spin arrangements of the 2D XY system. (a) and (b) depict the situation in the presence of vortices, and clearly there is a substantial disturbance. By contrast, there is less disturbance in (c) which illustrates the spin wave approximation.

infinity, implying that it would be very difficult to create a vortex in the thermodynamic limit. As we shall presently see, this statement will require some qualification.

Next consider the vortex-antivortex pair in figure 3. On account of a mutual cancellation effect, the spins at large distances are no longer severely disturbed, the disturbance being confined to the neighbourhood of the pair. As may be expected, the energy of the pair is finite, being given by

\[ E_{(\text{pair})} = 2E_c + 2\pi J \ln \left(\frac{r}{a_0}\right), \]

where \( r \) is the separation.

Going back to the single vortex, although its creation is energy-expensive, there is an entropy advantage associated with the fact that one can locate the vortex in various
Figure 3. Vortex-antivortex pair in the 2D XY system.

ways. This entropy is $2k_B \ln (R/a_0)$ so that the free energy is

$$F = E - TS = (\pi J - 2k_B T) \ln (R/a_0),$$

(7)

implying that for $T > T_c$ where

$$T_c = \pi J/2k_B,$$

(8)

free vortices may indeed exist even in the thermodynamic limit.

Based on the above arguments, KT proposed the following scenario: At low temperatures, only vortex pairs can exist. These are thermally generated and ride over a spin wave background. Since (bound) pairs do not cause serious perturbations at large distances (see figure 3), $C(r)$ is essentially that given by the spin wave approximation. Above $T_c$, the pairs dissociate since the system can now support free vortices. Technical considerations (to be amplified later) demand that the number of vortices be small. Even so, they drastically modify $C(r)$ (see also figure 2). As already noted, with such strong fluctuations present, the spin wave approximation does not apply. Further, like their counterparts in solids namely dislocations, vortices can move freely when subjected to suitable forces. The system as a whole therefore acquires flow properties and exhibits a concomitant loss of (spin) rigidity at $T_c$. This then is the central feature of the transition, the unbinding of pairs at $T_c$ and the related loss of the corresponding rigidity.

3.3 Topological order

Though conventional LRO is absent, KT suggest that one may differentiate between the phases on either side of $T_c$ via the concept of topological order. Briefly, topological order is an assessment of whether the topological defects are bound ("ordered") or unbound ("disordered"). To make such an assessment, one makes a large closed circuit in the system and measures the phase angle all along the boundary. The total change of
the phase angle will be determined by the number and strength of the singular points
enclosed by the path. In the high temperature phase (figure 4a), there will be a
preponderance of isolated singularities. The number enclosed will be proportional to
the area A of the contour C and the average total phase change will be proportional to
the length L of C. In the low-temperature phase (figure 4b), the average total phase
change will be determined by the number of pairs cut by the path, and will be
proportional to $(Ld)^{1/2}$ where $d$ is the mean separation of the pairs. Based on this
criterion, one can recognize a change in the topological order at $T_c$. It is worth noting
that a similar assessment by traversing a closed contour is also made in lattice gauge
theories (Kogut 1979).

3.4 2D melting—The KT view

We turn now to melting in two dimensions. Analogous to the spin wave approximation,
we have the harmonic (phonon) approximation. Introducing

$$\rho_G(R) \equiv \exp(iG \cdot R), \quad (9)$$

where $G$ denotes a reciprocal lattice vector, we may as usual define the correlation
function

$$C_G(R) = \langle \rho_G(R) \rho_G^*(0) \rangle, \quad (10)$$

which, for the 2D solid, varies as

$$\lim_{R \to \infty} C_G(R) \sim R^{-\eta_G(R)}, \quad (11)$$

Figure 4. Assessment of topological order is made by going round a simple closed contour
and measuring the overall change in phase angle. As discussed in text, the result depends on
whether the topological defects are unpaired (as in (a)) or are paired (as in (b)).
One can now calculate the structure factor

\[ S(q) = \langle |\rho(q)|^2 \rangle, \]

whereupon one finds that the familiar \( \delta \)-function Bragg singularities \( i.e \)

\[ S(q) \sim \delta(q - G) \]

become “smeared” into power law singularities (Jancovici 1967) \( i.e. \),

\[ S(q) \sim |q - G|^{-2 + \eta_G(T)} \]

One therefore has a “quasi” crystal which, incidentally, is the reason why one may still label the diffraction peaks with the set \( \{ G \} \). Well-defined peaks in \( S(q) \) can be observed provided \( \eta_G < 2 \).

Now dislocation-mediated melting is really an old concept in three dimensions, going back to the thirties (in a sense to Mott and Gurney 1939; for a review, see Cotterill 1980). With respect to two dimensions\( \ast \), \( \kappa_T \) note that at low temperatures, the system, though lacking crystallinity (see equation (13)), is a solid in that it is able to resist external shear. Besides phonons, the system also supports bound dislocation pairs which are thermally generated. The latter neither upset the power law decay of \( C_G(r) \) (see equation (11)) nor the rigidity. Above \( T_m \) the melting temperature the pairs break, and once free dislocations appear the system loses (elastic) rigidity, signalling a transition to the fluid state. The topological order can be assessed as earlier \( i.e. \) by making the Burgers circuit and measuring the failure in the path closure.

Regarding the details of the transition, one clearly needs a more refined analysis than contained in equations (5)—(8). The crucial feature in such an analysis is the screening of the pair interaction (6). The underlying physics is better appreciated by considering the analogous case of a 2D gas of charged particles\( \dagger \) with charges \( \pm q \), electrically neutral on the whole, and interacting \( \text{via} \) the logarithmic potential

\[ U(|r_i - r_j|) = -2q_iq_j \ln \left| \frac{r_i - r_j}{a_0} \right| + 2\mu, \quad r > a_0 \]

\[ = 0, \quad r < a_0. \]

Here \( q_i \) and \( r_j \) are the charge and position of the \( i \)th particle. \( 2\mu \) is the energy required to create a pair of particles of equal and opposite charge a distance \( a_0 \) apart. In 2D solids, the analogue of \( \mu \) is the dislocation core energy \( E_c \) (see also equation (5)). The quantity \( a_0 \) is an appropriate cut-off of the order of the particle diameter or, for a lattice, the lattice spacing. One expects that due to (14), opposite charges can link to form bound pairs, \( i.e. \) dipoles. For these pairs to unbind, \( U(r) \) must decrease, which can happen by screening.

Focusing attention on a particular dipole of size \( r \), the interaction between the two members constituting the dipole will be screened by (smaller) pairs lying within the

\( \ast \) For historical completeness it must be remarked that Feynman had independently proposed the idea of dislocation-assisted melting in two dimensions. Unfortunately, this work is unpublished. An outline of this theory is given by Elgin and Goodstein (1974) and by Dash (1978).

\( \dagger \) By now the reader would undoubtedly have noticed that there are a variety of 2D systems. For elucidating various features of the \( \kappa_T \) theory, we cite different examples as convenient!
range of the field. These pairs will in turn be screened by others lying in their respective
fields and so on. One thus has a scaling situation as in the case of critical fluctuations
present during a conventional second-order phase transition, and not surprisingly,
renormalization group techniques make their entry.

Owing to screening, (14) becomes modified to

$$U_{\text{eff}}(r) = \frac{2q^2 \ln (r/a_0)}{\varepsilon(r)},$$

(15)

where $\varepsilon(r)$ is the space-dependent dielectric function, and includes the polarizability of
pairs with separation less than $r$. Introducing $p(r)$ the probability of finding a pair with
separation $r$, one finds that $\varepsilon(r)$ and $p(r)$ are coupled as follows (Kosterlitz 1974;
Halperin 1979):

$$\varepsilon(r + dr) = \varepsilon(r) + 4\pi \times \text{pol. contributed by pairs in the}
r\text{and } (r + dr)
$$

$$= \varepsilon(r) + 4\pi \frac{r^2}{2T} 2\pi p(r) dr,$$

$$p(r + dr) = p(r) \exp \left[ -\frac{2}{rT \varepsilon(r)} \right] dr.$$

(16)

Here $r^2/2T$ is the polarizability of a pair with separation $r$ and $[2dr/\varepsilon(r)]$ is the work
done in increasing the separation of the pair by the amount $dr$. Using an iterative
procedure, one obtains from (16) the famous Kosterlitz recursion relations (Kosterlitz 1974).
Such relations can be written down for every $(2, 2)$ system i.e. the $XY$ model, 2D

crystal etc.

We turn to the outcome of the recursive relations in the specific case of 2D crystal.
Two crucial quantities are:

(i) $y = \exp (-E_c/k_B T)$, and

(ii) $K = \frac{4\mu(\mu + \lambda) a^2}{(2\mu + \lambda) k_B T},$

(17)

where $\mu$ and $\lambda$ are the Lamé constants and $a$ the lattice spacing. During the
renormalization calculations, one has to rescale the dislocation core diameter from $a_0$
to $a_0 \exp l$ causing $y$ and $K$ to become $l$-dependent. The recursive relations are of the
form (Kosterlitz 1974; Nelson 1978; Nelson and Halperin 1979; Young 1979):

$$\frac{dy(l)}{dl} = f_1(y(l), K(l))$$

$$\frac{dK^{-1}(l)}{dl} = f_2(y(l), K(l)).$$

(18)

The solution of these equations is too intricate to be discussed here. Of greater
interest are the renormalization group flows illustrated in figure 5. The different
trajectories effectively determine the system behaviour at different temperatures.

Let us follow a typical trajectory for $T < T_m$ from the starting line (see figure 5). The
flow takes us to the abscissa where $y(l)$ is zero and $K(l)$ has a finite value, which implies
that (i) the probability of finding a free dislocation is zero and (ii) the renormalized
stiffness is finite. The trajectory for \( T = T_m \) is the separatrix, and trajectories to its right flow away, implying that for \( T > T_m \) (i) free dislocations can be found and (ii) the system loses rigidity since \( K(l) \rightarrow 0 \). The system behaviour close to \( T_m \) may be carefully examined by zooming near the region where the separatrix curve meets the abscissa. The principal findings are:

(i) \[ C_G(R) \sim R^{-\eta_G(T)}, \quad T \rightarrow T_m^-, \]  
where \( \eta_G(T) \) is related to the renormalized Lamé constants \( \mu_R(T) \) and \( \lambda_R(T) \) by

\[ \eta_G(T) = k_B T \left| G \right| \frac{(3\mu_R + \lambda_R)}{4\pi \mu_R(2\mu_R + \lambda_R)}. \]

At \( T_m \), the exponent for the first Bragg point for a triangular lattice cannot exceed \((1/3)\).

(ii) \[ C_G(R) \sim \exp \left[ -\frac{R}{\xi_+(T)} \right], \quad T \rightarrow T_m^+, \]  
the correlation length \( \xi_+ \) diverging according to

\[ \xi_+ \sim \exp \left[ \frac{\text{const}}{(T - T_m)} \right], \quad \bar{v} = 0.36963. \]

(iii) The renormalized stiffness constant \( K_R(T) \) (see equation (14)) varies as

\[ K_R^{-1}(T \rightarrow T_m^-) \sim \frac{1}{16\pi} (1 - C |t|v), \]

where \( C \) is a positive, nonuniversal constant and \( t = (T - T_m)/T_m \). Above \( T_m \) the stiffness vanishes; \( K_R \) therefore abruptly jumps to zero from the value \( 16\pi \) at \( T = T_m \).

(iv) The specific heat has only an essential singularity. Physically this is related to the fact that at \( T_m \) the density of dislocation is rather low. There is therefore very little

---

**Figure 5.** Renormalization group flow in the KT theory. The quantities \( K \) and \( y \) are defined in (17). The line for \( T = T_m \) is the separatrix. The dislocation probability \( y(l) \) tends to zero in the region to the left of the incoming separatrix which is the crystalline phase. The trajectories to the right flow away to large \( y \)-values, signalling instability and the dissociation of dislocation pairs.
entropy associated with it. However above $T_m$ there is a small bump (Berker and Nelson 1979).

Next to the order parameter, the quantities of interest in a conventional phase transition are those which exhibit singularities. In the case of the $kT$ transition, there is no ordering in the conventional sense. The specific heat also is not very informative. The only meaningful signatures are the translational correlations $C_6(R)$ on either side of $T_m$ (see (19) and (21)), and the temperature dependence of the stiffness (see (23)), all of which are conveniently studied via x-ray and neutron scattering.

A few other related points must also be noted at this juncture. Firstly, the calculation of the screening via (16) is a self-consistent one, and for this reason the $kT$ theory is essentially a mean field type of theory. Secondly, while solving the recursive equations, the approximation is made that $y$ is small (i.e. defect density is small; see also remark (iv) above), implying that the core energy $E_c$ is large (see (17)). So the picture ($T > T_m$) is that of a dilute system of dislocations floating about in an otherwise harmonic solid. Should the cores overlap, the theory collapses. In a superfluid, the vortex core size is $\sim 1$ Å and overlap is not a serious problem. On the other hand, the core size of a dislocation could be several lattice spacings, and overlap becomes a real possibility (Kosterlitz 1980). One must therefore be ready for surprises. We shall revert to this question later. Finally, the transition is continuous; note especially the divergence of the correlation length (equation (22)), reminiscent of a conventional second-order phase transition. (However, a significant difference is that $\xi$ has an exponential rather than a power law behaviour, Stanley 1971.)

### 3.4 Role of orientational correlations

About five years after $kT$ advanced their theory, Halperin and Nelson (1978; see also Nelson and Halperin 1979) drew attention to an important aspect of the problem not considered by KT*. Nelson and Halperin (NH) argued that besides translational correlations, one must also consider orientational correlations, and that when a solid melts to become an isotropic liquid, it sheds both these correlations. NH therefore introduce an orientational order parameter $\psi$. Most of the focus so far has been on triangular lattices which are the most close packed structures in two dimensions. For this lattice, the bond orientational parameter is given by

$$\psi(r) = \exp[i6\theta(r)], \quad (24)$$

where $\theta(r)$ is the orientation relative to the bond between nearest neighbour atoms. $\theta$ is related to the displacement field $u(r)$ by

$$\theta(r) = \frac{1}{2} \left( \frac{\partial u_x(r)}{\partial x} - \frac{\partial u_y(r)}{\partial y} \right). \quad (25)$$

Analogous to (10), one now has a orientational correlation function

$$C_6(r) \equiv \langle \psi(0)\psi^*(r) \rangle. \quad (26)$$

Extending attention to orientational effects requires one now to consider new

* Taking into account a related paper by Young (1979), the melting theory is often referred to as KTHNY theory.
defects, namely disclinations. These are "rotational singularities", two examples of which are illustrated in figure 6. They are formed by taking a perfect triangular lattice and either removing or adding a wedge. The curvatures resulting thereof are significant (see, for example, Nelson 1983a) but we shall not consider them here. The interesting point about disclinations is that a dislocation can be viewed as resulting from the binding of a disclination pair, as illustrated in figure 6. In other words, a dislocation can be regarded as a disclination dipole.

Based on the above additional concepts, NH sketch a scenario for melting involving not only dislocations but disclinations as well. At low temperatures, the dislocations are paired, and, on account of the internal structure of the dislocation (see figure 6), the dislocation pair can be viewed as a disclination quadrupole. The harmonic phonons as usual disrupt positional LRO leading to a power law behaviour for the translational correlations as in (11), but, by contrast, the orientations exhibit LRO i.e.,

$$\lim_{r \to \infty} C_6(r) = |\langle \psi \rangle|,$$

(27)
a constant which is temperature-dependent.

Figure 6. (a) and (b) illustrate respectively five- and seven- fold rotational singularities (i.e. disclinations). For convenience, the atoms are suppressed and only the surrounding Wigner-Seitz cells are shown. (c) illustrates how a pair of disclinations can combine to form a dislocation. Here the atoms are shown along with their W-S cells.
As the temperature is gradually increased, a stage is reached \((T = T_m)\) where dislocations unbind. This, however, does not produce an isotropic liquid; instead the system enters an intermediate phase in which, \(C_6(r)\) decays exponentially as in (21) but rotational correlations, though asymptotically decaying, are still strong, being given by

\[
C_6(r) \sim r^{-\eta_6(T)}, \quad T \to T_m^+, \tag{28}
\]

where

\[
\eta_6(T) = 18 k_B T / \pi K_A(T). \tag{29}
\]

The quantity \(K_A(T)\) called the Frank constant is related to the so-called bending energy of orientational fluctuations as

\[
H_A = \frac{1}{2} K_A \int d^2 r (\nabla \theta)^2. \tag{30}
\]

In effect, \(K_A\) is a measure of the “angular stiffness” of the system. Below \(T_m\), \(K_A = \infty\) indicative of perfect angular correlations but above \(T_m\), there is a temperature dependence given by

\[
K_A(T) \sim \xi_+^2(T), \quad T \to T_m^+, \tag{31}
\]

where \(\xi_+(T)\) is the same as in (22). This divergence is of course linked to the appearance of long range orientational order below \(T_m\) (see (27)).

The phase with properties as in (21), (28) and (31) is like a liquid crystal in that it has greater orientational order (power law decay) than translational order (exponential decay). This phase is frequently referred to as a hexatic phase.

When heated, the hexatic phase in turn undergoes a transition at a temperature \(T_i\), becoming an isotropic liquid. This occurs via the unbinding of the disclination pair, causing a loss of the orientational correlations present in the hexatic phase. Thus, in the NH picture, the free topological defects in the isotropic liquid are not the dislocations but the disclinations. As noted by NH, the transition at \(T_i\) belongs to the same universality class as the two-dimensional superfluid and the \(XY\) model, whence,

\[
\eta_6 \to \frac{1}{4} \quad \text{as} \quad T \to T_i^- \tag{32}
\]

The renormalized Frank constant jumps discontinuously to zero at \(T_i\) from the value \(72 k_B T / \pi\) to which it decreases as \(T \to T_i^-\). Above \(T_i\),

\[
C_6(r) \sim \exp \left( -r / \xi_6(T) \right) \tag{33}
\]

with

\[
\xi_6(T) \sim \exp \left[ \frac{b}{(T - T_i)^{1/2}} \right]. \tag{34}
\]

As usual, there is only an essential singularity in the specific heat. In short, at \(T_i\) there is again a crossover from a power law to exponential decay, but this time, of the orientational correlation function. Accompanying this change is a vanishing of the orientational stiffness.

Figure 7 paraphrases the features of the original \(KT\) theory and the subsequent \(NH\) modification. In passing we note that although Halperin and Nelson (1978) also predicted continuous transitions, they were cautious enough to observe that they “cannot rule out other mechanisms for melting, perhaps leading to a first-order transition”.

Melting in two dimensions
Laboratory experiments on 2D melting are frequently performed on layer systems supported by a substrate. NH have examined the role played by the substrate by adding an orienting field to the defect Hamiltonian, and find the following: If the substrate can be regarded as smooth, then the considerations of figure 7 continue to remain valid. On the other hand, if the periodic nature of the perturbing potential from the substrate is significant, then one must examine whether the periodicity of the substrate is commensurate or not with that of the layer above (see also figure 8). Of interest is the case where the substrate potential is weak and incommensurate. In this situation, there is only one transition, namely, that at $T_m$; the other one at $T_i$ is washed out (for a triangular lattice).

A second question relates to phase diagrams. In condensed matter, we are used to diagrams as in figures 9a and 9b. One would like to know how these are modified in the light of the $kT$NY theory. The outcome is sketched in figure 9c and 9d, and will become relevant later.

The prelude has been somewhat lengthy but is indicative of the fact that 2D systems are a theorist’s paradise! As we have seen, topological defects in 2D are point defects. It turns out that the statistical mechanics of these is a lot simpler than that of line defects (with which one has to deal in 3D). Hence the great popularity of 2D systems with theoreticians. We turn now to the various attempts made to check the $kT$NY theory.
Melting in two dimensions

Figure 8. Schematic illustration of commensurate and incommensurate coverage of graphite by argon atoms. The Ar-Ar Lennard-Jones potential is sketched at the bottom. (After Mc Tague et al. 1980b.)
4. Experimental studies

4.1 Laboratory experiments

There are a number of candidate systems like liquid crystal films, electron films on liquid helium, etc which are suitable for experimentally checking out the \textit{kthny} theory. Many of these have indeed been investigated (Sinha 1980) but the most detailed studies so far have been on rare gas films absorbed on graphite.

The basal i.e (0001) plane of graphite is a good substrate for forming monolayers by physisorption since it presents a relatively smooth potential to the layer absorbed. Unfortunately, large single crystals of graphite are not available but this has not deterred experimentalists. Instead they have skilfully exploited various exfoliated graphite products. Basically all of these are built up of small flakes a few \( \mu \text{m} \) broad and \( \sim 100-300 \text{ Å} \) thick. The useful size is characterized by the coherence length \( L \) (see figure 10). The \( c \) axes of these flakes are all partially oriented about the normal to the sheet but the azimuthal orientation about the \( c \) axis is random. Various types of exfoliated graphite like Grafoil, UCAR-XYZ and Papyex are now commercially available. Very popular (especially in the past), Grafoil has \( L \sim 200 \text{ Å}, \Delta \theta \sim 25^\circ \) (see figure 10) and a specific area for adsorption \( \sim 30 \text{ m}^2/\text{g} \). The samples (for scattering experiments) typically consist of a pile of discs cut from sheets of exfoliated material. Samples with several grams of Grafoil are not uncommon. Very recently, “single crystals” of exfoliated graphite have been prepared (Clarke \textit{et al} to be published; quoted in Rosenbaum \textit{et al} (1983)) with \( L \sim 400 \text{ Å} \) and \( \Delta \theta \sim 3^\circ \).

Although notionally the graphite substrate may be deemed to exert only a weak influence, the role of the substrate cannot be entirely overlooked. As already noted in figure 8, the adsorbed layer could be commensurate* as well as incommensurate with the substrate. Our interest is in the latter situation.

![Figure 10. Schematic drawing of the Grafoil sheet. \( L \) indicates the typical coherence length of the adsorbed layers while \( P(\theta) \) describes the distribution arising from the non-parallelism of the substrate surfaces (After Nielsen \textit{et al} 1980).](image)

* It is interesting that if the adsorbed atoms are in registry with the substrate, \( \langle u^2 \rangle \) does not diverge (Kosterlitz 1980).
Monolayers of Ar, Kr and Xe on graphite have been studied using both neutron and x-ray scattering (Thorel et al. 1976; Taub et al. 1977; Horn et al. 1978; Stephens et al. 1979; Hammond et al. 1980; Heiney et al. 1982; McTague et al. 1982; Heiney et al. 1983) and the opinion has emerged that Xe on graphite is by far the best representative of 2D field-free system (McTague et al. 1982). Accordingly we shall examine the recent high resolution results of Heiney et al. (1983) obtained at the Stanford Synchrotron Radiation Laboratory (SSRL). These experiments are a sequel to an earlier series done at MIT using a rotating anode source but with somewhat lower resolution.

Heiney et al. have criss-crossed the phase diagram along various trajectories (including constant temperature ones) in order to check out different possibilities. Figure 11 shows representative profiles observed in one of these runs. The asymmetric lineshapes are due to excess high q scattering arising from substrate azimuthal disorder. Around 152 K the lineshape changes, going over to a Lorentzian characteristic of exponential decay of the correlation function. Heiney et al. reckon that a transition

Figure 11. Diffraction profiles of the (1, 0) spot of xenon adsorbed on graphite. The scans have been made in the narrow region between 151.6 K to 160 K (After Heiney et al. 1983).
occurs at 152 K. The question whether it is of first order or continuous has been examined quite carefully. Firstly, no hysteresis is observed. Secondly, as the transition temperature is approached from below, the exponent obtained from the lineshape analysis (Dutta and Sinha 1981) has the trend shown in figure 12. At the melting temperature it is found that 0.27 ≤ η (melting) ≤ 0.42, reasonably consistent with KTHNY prediction 0.25 ≤ η(T_m) ≤ 0.33. Above 152 K where the line shape is consistent with (22), analysis yields v = 0.4 (see also figure 12).

The evidence cited above no doubt strongly favours the KTHNY theory. Nevertheless Heiney et al have carefully examined possible consistency of their data with first-order transition. If the transition is indeed of first-order then near the transition temperature one must encounter a coexistence patch. Assuming that the solid material is converted into a liquid according to a lever law, they have analyzed their data. From the values of χ² obtained they conclude that the observed melting is consistent with a continuous transition although a possible two-phase coexistence in a very narrow range between 151.6 and 151.95 K cannot be ruled out.

Recently, Rosebaum et al (1983) studied the melting of monolayer xenon on the surface of exfoliated single crystals, using x-ray scattering. According to the NH theory,
in the solid one must get six \((1, 0)\) Bragg spots, which, owing to the lack of LRO, will have some width both in the radial as well as angular directions with probably more in the former. On melting into the hexatic phase, the six spots still survive\(^*\) with, however, enhanced width in both directions. As temperature is increased, the changes in the orientational correlations will be more rapid than the corresponding ones in the positional correlation. Consequently the angular width will increase more rapidly compared to the radial width until at \(T_x\) the spots merge to form a liquid ring. Such a behaviour (see figure 13) is in fact seen in experiments. Rosenbaum et al categorically rule out from the observed lineshapes, the possibility of a liquid-solid coexistence. Nor do they believe that the observed spots could be due to an isotropic liquid subjected to a substrate field. The field strength required would be three orders of magnitude too high. They are confident that their experiment offers evidence for melting of the 2D solid into an orientationally ordered liquid.

Besides adsorbed layers, melting of liquid crystal films also have been investigated to test the validity of the \(k\)thy theory. An important feature of these experiments is the use of freely suspended films, the preparation of which is schematically illustrated in figure 14. While a monolayer film has yet to be realized, films with as low as two layers have been successfully used.

Moncton and coworkers (Moncton and Pindak 1979, 1980; Moncton et al 1982) have been particularly active in the study of x-ray diffraction from such films. An early experiment (Moncton and Pindak 1980) on a material called 40-8 (butyloxybenzylidene octylanline) seemed to suggest that the melting was continuous with \(\eta\) at melting in the range specified by the \(k\)thy theory. However, films thinner than 4 layers could not be studied owing to film rupture problems. Subsequently, they switched to another material 14S5 (Moncton et al 1982) which was amenable to the formation of films with as low as two layers. It was found that the film with two molecular layers melted by a

\[\text{Figure 13. Schematic drawing to illustrate the outcome of the experiment of Rosenbaum et al (1983). Sketched here is the related variation of radial and angular widths of the (1, 0) diffraction spot as temperature is increased. As } T \text{ becomes larger, the angular width increases more rapidly causing the six spots to merge into a ring.}\]

\(^*\) Technically, in the hexatic phase the mean square angular fluctuations \(\langle \delta \theta^2 \rangle \rightarrow \infty\) and x-ray scattering should give a ring. However, a weak substrate field is enough to restore the six spots. On the other hand, if the orientational correlations are not strong (i.e. are liquid like as in (33)), then a very strong substrate field would be needed to produce the spots.
Table 1. Summary of papers dealing with computer simulation experiments.

<table>
<thead>
<tr>
<th>Reference</th>
<th>System considered</th>
<th>Results</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morf (1979)</td>
<td>780 electrons. Molecular dynamics.</td>
<td>Shear modulus shows a temperature dependence similar to what is indicated by $kT$.</td>
<td></td>
</tr>
<tr>
<td>Broughton et al (1982)</td>
<td>780 particles, $1/r^{12}$ potential. Molecular dynamics. Isothermal scan.</td>
<td>First-order transition. $K$ versus $P$ indicates a value at melting close to $16\pi$ predicted by $kT$.</td>
<td>Value of $K$ of $16\pi$ at melting is surprising since transition is first order. A similar result has been obtained by Abraham.</td>
</tr>
<tr>
<td>McTague et al 1980a</td>
<td>256 and 2500 $r^{-6}$ particles. Monte-Carlo. Isochore traversed.</td>
<td>Defect structures examined and grain boundary loops found. Unable to confirm if transition is continuous.</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Methodology</td>
<td>Results</td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Toxvaerd (1980)</td>
<td>256 and 3600 LJ particles. Constant temperature scan.</td>
<td>First order transition seen for both systems. Results for high density systems consistent with first-order transition. In the low density system a discontinuity in K is observed close to 16 π.</td>
<td></td>
</tr>
<tr>
<td>Abraham (1981b)</td>
<td>576 xenon atoms on graphite. Isothermal and constant pressure scans.</td>
<td>Results compared with x-ray experiments. First-order transition observed in contrast to lab. experiments.</td>
<td></td>
</tr>
<tr>
<td>Koch and Abraham (1983)</td>
<td>Xenon films on graphite of thickness greater than one monolayer simulated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abraham (1984)</td>
<td></td>
<td>Apparently Frenkel and McTague misinterpreted the two-phase region as the hexatic phase. Authors comment that increasing the size weakens the first-order transition.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Special hardware processor used for the simulation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abraham feels the kT melting temperature is an upper limit for the stability of the metastable (i.e. superplated) 2D solid. This study was to see if substrate had any effect. Computational box as in figure 21.</td>
<td></td>
</tr>
</tbody>
</table>
single abrupt transition with hysteresis and no critical behaviour i.e. by a first-order transition. These results are also in accordance with the findings of Bishop et al (1982) in a related study in which shear mechanical measurements were made.

We have confined our survey to the most important of the several experiments done so far. While the studies on adsorbed layers seem to favour the KTHNY theory, the melting of liquid crystal films seem to point to a first order transition. Since both these classes of experiments are open to their own objections, many have sought an answer via computer simulation to which we now turn our attention.

4.2 Simulation experiments

Computer experiments on phase transitions have a long history going back to the pioneering work of Alder and Wainwright (1962) who studied hard disc systems. The experiments we review use either the Monte-Carlo approach (Metropolis et al 1953) or the molecular dynamics approach (Rahman 1964). A large number of papers have been published, of which table 1 offers a partial summary. Given this large volume of work, we can indicate only the highlights.

4.2a Advantages—An important advantage of the simulation experiment is that it enables one to follow events at the level of individual atoms. To appreciate this, we refer to figure 15. Here 15a shows an instantaneous snapshot of the system. The coordination number of each atom in this assembly can now be analyzed via a Dirichlet/Wigner-Seitz construction (see figure 6), and one can identify those with anomalous coordination i.e. those not having the normal quota of six neighbours. One can now suppress the “normal” atoms and project only the defect structure as done in figure 15b. Such studies have not only shed much light on how defects evolve and multiply, but more significant, have called attention to the possible role of grain boundaries in the melting process.

A second useful pointer to emerge is the superiority of the constant pressure and isothermal scans as opposed to traversal along an isochore (constant density scan). In the former, the distinction between a first order and a continuous transition is more clear-cut (as we shall presently see). In an isochore traversal on the other hand, one could strike a coexistence patch, which, besides making the transition appear
Figure 15. (a) shows a typical configuration of atoms during the simulation. + and − denote respectively seven- and five-fold coordinations. In (b) is projected the defect structure alone (After McTague et al 1980a).

Figure 16. Equilibrium density and enthalpy per atom as a function of temperature for a 256 LJ atom system at $P^* = 0.01$. The sharp breaks are characteristic of first-order transition (After Abraham 1980).
continuous, could also give erroneous indications about the presence of a hexatic phase.

Yet another merit of simulation is the ability to follow the trajectories of individual atoms over a fairly large number of steps. We shall shortly consider examples of such plots and their utility.

4.2b Some results—In a typical simulation experiment, one considers an assembly of particles interacting via a suitable potential. The particles are confined to a planar cell and suitable boundary conditions are imposed (usually periodic boundary conditions; exceptions however have been made: Hansen et al 1979; Abraham 1981b). The thermodynamic conditions like pressure, temperature etc are specified and the simulation is executed in a series of consecutive steps. Initially, a large number of steps must be gone through (with suitable “steering”) so that equilibrium is attained corresponding to the desired conditions. Near a transition, special care is necessary since equilibration will require many more steps than usual. After attainment of equilibrium, the system is run through more steps, during which “measurements” are made. In a constant pressure run, for example, one would monitor the density and enthalpy. Besides evaluating thermodynamic quantities, correlation functions are also sometimes computed. In addition, defect patterns, trajectory plots etc are obtained as required. This sequence is then repeated for a new set of initial conditions and in this way a whole scan is made.

We consider now some representative results. Figure 16 shows the equilibrium density and enthalpy per atom as a function of temperature for a 256 Lennard-Jones \( V(r) = 4e \left\{ (\sigma/r)^{12} - (\sigma/r^6) \right\} \) atom system at a fixed pressure. Abraham (1980) who did this work, carried out such studies at three reduced pressures \( P* = \frac{P_0}{e} = 0.01, 0.05 \) and 1. The first two pressures were intentionally chosen low since NH had speculated that the melting could be first order at high pressures but continuous at low pressures (see also figure 9). Abraham however finds that melting is always a discontinuous process as in figure 16.

Figure 17 shows results for isothermal scan obtained by Toxvaerd (1981) for two Lennard-Jones systems with 256 and 3600 particles respectively. Both exhibit Van der Waal loops characteristic of a first order transition, the loop being smaller for the larger system. Whether the transition will continue to be first order in the thermodynamic limit is difficult to say but after investigating the size dependence from 256 to 3600 particles, Toxvaerd seems to feel that the transition will continue to remain weakly first order. Broughton et al (1982) investigated a system of 780 particles interacting via a

![Figure 17. Pressure versus density at constant temperature for two LJ systems. The solid circles are for 3600 particles and the open circles are for 256 particles (After Toxvaerd 1981).](image-url)
purely repulsive $r^{-12}$ potential. They too find a weak, first-order transition. These authors also studied a system with a single vacancy which gave them the same loop during both forward and reverse scans, unlike the perfect system. Both Toxvaerd and Broughton et al have detected grain boundaries from defect structure analysis (recall figure 15). They also cautiously note that periodic boundary conditions could stabilize the perfect solid, inhibiting the formation of vacancies and interstitials. In turn this could inhibit dislocation climb and modify the character of the transition. However, taking all factors into account, the overall assessment of both authors is that the melting transition is not only weakly first order but just preempts the $kT$ transition.

Special mention must be made of the work of Bakker et al (1982, 1984) who have built a dedicated hardware processor enabling them to make a quantum jump. Instead of dealing with 512 or 1024 particles as usual, they are able to study the molecular dynamics of 10864 particles. Their device had the speed of a super computer, achieving in 750 hours what would have taken 4000 hours in Amdahl 470 V/7-B. One expects this approach will set a new trend. Incidentally, Bakker et al (1984) have studied a Lennard-Jones system along the isochore $\rho^* = \rho \sigma^2 = 0.88$, very close to that traversed earlier by Frenkel and McTague (1979) and Toxvaerd (1980). Whereas Frenkel and McTague interpreted their data as being indicative of two continuous transitions with an intermediate hexatic phase, Bakker et al, in agreement with Toxvaerd, conclude that melting is in fact first order, and that the intermediate phase is really a coexistence region.

This is a convenient juncture to introduce trajectory plots, a good example of which is given in figure 18. The plots here pertain to a 256 atom Lennard-Jones system, and were obtained during a scan along an isochore (Abraham 1981a). They show four consecutive sequences, each generated from $8 \times 10^5$ Monte-Carlo steps. In each plot one can see a crystalline as well as a liquid-like region, indicative of coexistence. Further, the melted region appears to be mobile with passage of time, suggestive of a small interfacial energy between the liquid and the solid. Such a coexistence could well mimic a hexatic phase, even if the phase were absent.

Figure 18. A consecutive sequence for four trajectory plots for a 256 LJ atom system. Each plot is generated from $8 \times 10^5$ consecutive Monte-Carlo steps, with the system held at fixed density and temperature. Observe the coexistence of solid and liquid regions (After Abraham 1981a).
The simulations discussed thus far deal with systems having no physical boundaries. Abraham (1981a, b) considered the role of surfaces since there is some evidence for surface premelting in the case of 3D crystals (Broughton and Woodcock 1978). He finds (see also figure 19) that a 2D solid with two surfaces is a stable crystal with premelted surface at reduced temperature $T^* = 0.40$ ($T^* = k_B T/\varepsilon$) and a stable liquid at $T^* = 0.42$. An independent free energy analysis of the 2D Lennard-Jones system yields a thermodynamic melting temperature of $T^* = 0.415$ (Barker et al 1981). Abraham further finds that a solid subjected to periodic boundary conditions melts at $T^* \sim 0.45$ with a discontinuous jump in density (i.e. via first order). From this Abraham concludes that a solid subjected to periodic boundary conditions superheats well beyond the thermodynamic melting temperature. As he remarks (Abraham 1981a) the melting temperature of the surfaceless solid "corresponds to an upper limit for the stability of the metastable 2D solid". Associated with the transition at $T^* = 0.45$, there is the stiffness dependence shown in figure 20. At the transition temperature, the measured value of $K$ is lower than $16\pi$ "because of defect formation in the solid constrained to remain at solid density". However, if $K$ values obtained for $T^* < 0.45$ are extrapolated, they yield the solid curve shown in figure 20. $K$ thus appears to jump discontinuously to zero at the transition from the universal value of $16\pi$. This of course emerges from the renormalization group analysis of $kT$. Why it should do so likewise in a first-order transition is not clear.

We have already alluded to some of the doubts that can be cast on the experiments on adsorbed films. One of these relates to the possible role of the second layer (which is frequently present). To make an assessment, Abraham (1983; 1984; see also Koch and Abraham 1983) simulated xenon films on graphite, the geometry of the computational box being as in figure 21. The simulations covered also the conditions of the experiments of Heiney et al (1982) and Rosenbaum et al (1983). Abraham's main conclusion is that the transition still remains first order, at variance with the findings of the x-ray experiments. Imperfections in the laboratory graphite substrate could be a possible source of the discrepancy.

Figure 19. Trajectory plots of a 2D 512 LJ atom system with two free surfaces. The two temperatures investigated are on either side of the thermodynamic melting point of 0.415. (After Abraham 1981a).
Melting in two dimensions

4.2c Some doubts—Though elegant, simulation also is not free from objections, principal among which are the following (Broughton et al 1982):

(i) Typical sample sizes of 500–1000 atoms are too small to be representative of thermodynamic specimens.

(ii) Even the largest computer simulation run corresponds to a very short time on a lab scale. Achievement of equilibrium near a transition is therefore a sensitive question, especially if critical slowing down is a possibility.

(iii) Periodic boundary conditions could modify the apparent nature of the transition.
(iv) Simulation without an adequate number of built-in vacancies and interstitials might not produce a sufficient amount of dislocation climb which is necessary for a proper test of the $kT$ theory.

There are a few other supplementary queries that may be raised. It is fair to say that those engaged in simulation have given serious consideration to all these objections in carrying out their experiments and in analyzing the results. Some observers are however not still convinced.

4.2. Simulation using defect Hamiltonian—We turn now to an entirely different kind of simulation due to Saito (1982a, b) who, instead of working with atoms, used a defect Hamiltonian defined on a triangular mesh with periodic boundary conditions. In the thermodynamic limit, the defect Hamiltonian goes over to that of $kT$ i.e. of continuum elasticity theory (as it should). Saito uses the Monte-Carlo method. A pair of nearest neighbour sites are chosen at random from the mesh. If both sites are empty, one tries to create a pair of dislocations with opposing Burgers vectors. If both sites are occupied with dislocations of opposite Burgers vectors, an annihilation is attempted; otherwise an exchange is tried. The trial is accepted according to the usual Boltzmann weight etc.

Saito finds that if $E_c$ the dislocation core energy is large, the melting transition is due to dislocation unbinding and is in accord with the $kT$ theory. For a small core energy, however, the melting is of first order, being caused by the nucleation of grain boundary loops. It is natural to speculate whether a system of atoms will show similar dichotomy depending on whether the inter-atomic potential favours a large $E_c$ or not. In collaboration with V. Sridhar and B. Chakraborty of our Centre, we are currently examining this question via simulation. Meanwhile, we learn that Van Himbergen (1984) has carried out a similar study for the ATT model and finds both continuous as first order transition, depending on the shape of the nearest neighbour interaction. One therefore awaits Sridhar’s results with interest.

5. Other mechanisms and viewpoints

In view of the doubts concerning the validity of the $kT$ hypothesis, several alternate viewpoints have recently been proposed. Guided by the results of several simulation studies and the work of Fisher et al (1979) who noted that grain boundaries could appear before $T_m$ is reached, Chui (1982, 1983) constructed a grain boundary theory of melting. Chui models the grain boundary as an array of dislocations. A pair of dislocations with Burger’s vectors $\mathbf{b}$, $\mathbf{b}'$ interact with each other via a potential of the form

$$V = -\frac{K}{4\pi} \left( \mathbf{b} \cdot \mathbf{b}' \ln |\mathbf{r} - \mathbf{r}'| - \frac{[\mathbf{b} \cdot (\mathbf{r} - \mathbf{r}')] [\mathbf{b}' \cdot (\mathbf{r} - \mathbf{r}')] }{|\mathbf{r} - \mathbf{r}'|^2} \right),$$

where $K$ is as in (17). The first term in (35) should be familiar to us from the corresponding term of (14). The second term of (35) represents the anisotropic part of the interaction. (The continuum Hamiltonian of $NH$, for example, will have a third term related to the core energy $E_c$. That, however, is not pertinent here.) A grain boundary configuration appropriate to a triangular lattice was next considered, and the effective potential between parallel grain boundaries at a distance $Z$ apart was computed using
Melting in two dimensions

(35). In turn this formed the input for a free energy analysis of phase stability. The conclusions drawn by Chui are:

(i) grain boundaries do make an appearance before \( T_m \),
(ii) a strongly first-order transition results if the grain boundaries are coupled to density change or bound dislocations pairs,
(iii) the transition becomes weakly first order for a core energy \( < 2.84 T_m \).

In short, grain boundary melting preempts the \( kT \) transition in Chui’s theory.

Ramakrishnan (1982, 1984) following his earlier work with Yussouff (1979), approaches the transition from the liquid side i.e as a freezing problem. According to this view, the solid emerges from the liquid primarily due to the dominance of a density wave of crystalline periodicity and symmetry. The wavevector \( G \) of this wave is identified with the first peak in \( S(q) \) the liquid structure factor. From the quantitative point of view, the theory considers the solid near melting as a calculable perturbation on the liquid and shows that their free-energy difference is determined by two parameters \( \mu(G) \) and \( c(G) \). Of these \( \mu(G) \) is defined by

\[
\rho(r) = \rho_0 (1 + \eta + \mu(G) \exp iG \cdot \gamma),
\]

where \( \rho(r) \) is the density. \( c(G) \) is the Fourier transform of the fluid correlation function and is related to the structure factor as

\[
S(G) = (1 - c(G))^{-1}.
\]

A self-consistent relation for \( \rho(r) \) is derived involving also the potential \( V(r) \), from which the free energy difference is computed. The theory predicts a first-order transition at a temperature \( T_f \).

Ramakrishnan (1982) also considers the implication of \( T_f \) being greater or less than the \( kT \) transition temperature \( T_m \). Of course if \( T_f < T_m \), then the \( kT \) mechanism has no chance. On the other hand if \( T_m < T_f \), it could well happen that owing to lack of sufficient run time, computer samples continue to appear as a solid and melt only when \( T_f \) is attained (via first order of course). Ramakrishnan therefore advocates suitable experiments on spatial correlation to ascertain whether \( T_f < T_m \) or \( T_f > T_m \) in systems where \( T_f \) is clearly identified.

In a series of high-powered (!) papers, Kleinert (1983a, b and references therein) has reexamined the \( kT \) mechanism, with particular reference to the role of the core energy term in the Hamiltonian for an assembly of defects. In the \( kT \) theory, the core energy term has a form \( \sim E_c \sum_i (b_i^2) \) with the further stipulation \( \sum_i b_i = 0 \). In the Coulomb model mentioned in §3.4, the corresponding expressions would be \( \sum_i q_i^2 \) with \( \sum_i q_i = 0 \) (charge neutrality).

According to Kleinert, a single disclination is really a point singularity at the end of a string of dislocations, rather like the “Dirac string” of the magnetic monopole. Kleinert’s view is that a core energy term of the \( kT \) type is inappropriate since it implies an infinite energy for the dislocation string, automatically ruling out dislocation pile up and a concomitant first-order transition. By following a route earlier established by Villain (1975) \textit{vis-a-vis} the 2D \( XY \) model, Kleinert derives an expression for the partition function which, unlike that of \( kT \), is free from this “objection”. An interesting feature of Kleinert’s work is the use of duality concept, already familiar in lattice gauge
theory (Kogut 1979), spin glass (Fradkin et al 1978) etc. The net outcome of Kleinert's analysis is that melting in two dimensions is a first-order process. Very recently, Janke and Kleinert (1984) have backed up this conclusion with a Monte-Carlo simulation.

6. Summary and concluding remarks

Despite the early start given by Peierls (1935) and Landau (1937) on the theoretical side and the pioneering experimental work of Langmuir (1938) on surfaces, 2D solids did not receive much attention for a long time, possibly because the departure from true crystallinity was almost notional. However, the deep enquiry in the sixties concerning the nature of ordering in 2D systems revived interest, leading to the novel proposal by Kosterlitz and Thouless (1973) that phase transitions are possible without being accompanied by ordering in the usual sense.

Several systems (superfluid films, electrons on liquid helium etc) have been experimented upon to test the validity of the kT idea (see for example, Sinha 1980). As far as melting in 2D is concerned, the important question is whether it is prompted by long wavelength or short wavelength fluctuations. If it is the former, then the transition is not only continuous but has several unusual features. A crucial factor is the role played by topological defects, in particular the dissociation of bound pairs. On the other hand if melting occurs mainly due to a pile-up of disruption of local arrangements i.e. due to short wavelength fluctuations (which is facilitated by core overlap—see remarks in §3.3), then the transition is of first order as in 3D.

We have presented a sampling of the more detailed (lab) experiments on melting carried out so far. Most of these have been on adsorbed films of rare gas atoms, and the results seem to favour the KTHNY theory. The results on self-supporting films of liquid crystals seem however to favour a first-order transition.

In contrast to the lab experiments, the computer experiments do not have problems of specimen preparation and therefore have been quite popular. The evidence from such simulations appears to be in favour of a first-order transition. However, as in the case of the lab experiments, the results are not totally objection-free. For some, no matter how detailed the simulation, these experiments come nowhere near to real life on account of the various factors discussed earlier. On the other hand, the practitioners of simulation by and large seem convinced that the transition is of first order, although some are more cautious than others in articulating their opinions. In passing, it is interesting to note that Abraham (1984) in his simulation studies is able to reinterpret the x-ray results of Heiney et al (1983) and of Rosenbaum et al (1983) in support of a first order transition, although the experimenters themselves feel their results point to a continuous transition.

At the moment there is lull of sorts, with each “camp” holding on to its own view

---

* See also Jose et al (1977) and Nelson (1978).

* Abraham (1980), for example, estimates that for a correlation loss ~ 6 A, the 2D crystal must have a size ~ 10^{10} light years! It is remarkable, however, that diffraction experiments reveal via line shapes even these minute differences.

1 Within the simulation camp itself there has been some cross fire! See Toxvaerd (1984) and Abraham and Koch (1984).
point. Resumption of vigorous activity is, however, very much on the cards especially on the lab experiments side, with accent on eliminating the various objections raised in the past. Besides the ever-popular diffraction studies, other supporting investigations may also be anticipated, e.g. on shear properties. Hopefully these will settle the issue once and for all in an unambiguous manner. Meanwhile one should not lose sight of the finding of Saito (1982a, b) that one can get either a first order or a continuous transition, depending on the value of $E_c$, the core energy. There is also the related work on the XY model (Van Himbergen 1984) where this dichotomy has been analyzed. It could well happen that there are some 2D lattices which melt via a $\kappa T$ mechanism and others which do not!

It would indeed be a pity if in the final analysis it is established that 2D melting does not occur by a $\kappa T$ process. If so, it would be a narrow miss, for the transition appears to be on the edge of being a $\kappa T$ type, as indicated by many of the experiments done so far. In a way this underscores how sensitive the outcome of a given experiment is to “disturbances” (e.g. imperfections in the substrate).

Even if it turns out that 2D melting is first order, the $\kappa T$ paper would not be without significance, having exerted a strong influence on the study of phase transitions in 2D. Of special interest is the influence exerted “across the border” on particle physics. This is not altogether surprising, given the parallels between statistical mechanics and field theory (Kogut 1979), and particle theorists’ interest in topological objects (Coleman 1977 and Rajaraman 1982). Thus, whatever be the final outcome as regards melting, the $\kappa T H N Y$ and all related papers have been beneficial in their own way to the progress of physics, having made a beginning concerning the statistical mechanics of topological defects.

7. Other reviews

In view of its great popularity, there naturally have been several reviews on the $\kappa T$ transition from time to time. It is pertinent therefore to call attention to these, and place our own in the proper perspective.

In one of the early reviews, Kosterlitz and Thouless (1978) emphasize superfluidity, understandable in view of the forum. A subsequent survey entitled “ordering in two dimensions” by Kosterlitz (1980) is broadbased but the accent is almost entirely on theory. Halperin has reviewed the subject in his Kyoto lectures (Halperin 1979) and in the Les Housches School (Halperin 1981). The discussions are extensive and again the emphasis in on theory. Barber (1980) does not restrict himself to the $\kappa T$ problem but covers instead the whole gamut of $O(n)$ models in 2D. In the process, he indicated linkages to the problem of quark confinement. Nelson (1979) has several useful and interesting comments on the $XY$ model and its various gauge generalizations. Young

* For example, Nelson (1983b) remarks “I doubt that the simulation techniques used by Abraham et al could reproduce the apparently continuous melting transition of incommensurate xenon on graphite observed recently via precision x-ray diffraction by Heiney et al.”. Abraham (1984) on the other hand holds, “From our present study the status is clear for the high-temperature xenon film on graphite . . . . We observe first-order melting of xenon . . . .”.

P.— 29
(1980) gives a simple and crisp introduction to the central idea of the $kT$ theory. Abraham (1981a) has a very readable article on simulation experiments but it is a paraphrase of his own extensive work. Brinkman et al (1982) offer a brief but popular survey of the 2D melting question. Mention should also be made of the conference volume edited by Sinha (1980) which gives a good overview of the then state of the art relating to phase transitions in 2D. Our paper restricts to the melting problem alone, but goes into some detail both on the conceptual as well as on the experimental side. The discussion of the experimental situation is reasonably uptodate. The present review should therefore be a useful complement to the various earlier surveys.

8. Dedication and acknowledgements

It is a pleasure to dedicate this article to Dr. R. Ramanna on the occasion of his sixtieth birthday for his many contributions, and especially his vigorous championing of basic research.

We are grateful to Dr. B. Chakraborty and Shri V. Sridhar for discussions, to Shri P. Subba Rao for typing assistance and to Shri P. Harikumar for help with the graphics. Above all, we are indebted to the many authors whose work we have freely drawn upon. One of the authors (GV) would like to thank the Jawaharlal Nehru Memorial Fund for the award of a Fellowship facilitating a concentration on the topological aspects of defects.

References

Abraham F F 1980 in Ordering in two dimensions (ed.) S K Sinha (Amsterdam: North Holland) p 155
Chui S T 1982 in Melting, localization, and chaos (eds.) R K Kalia and P Vashista (Amsterdam: North-Holland)
Cotterill R M J 1980 in Ordering in strongly fluctuating condensed matter systems (ed.) T Riste (New York: Plenum) p 261
Melting in two dimensions

Kleinert H 1983a Phys. Lett. A95 381
Kogut J B 1979 Rev. Mod. Phys. 51 659
Kosterlitz J M 1974 J. Phys. C7 1046
Langmuir I 1938 J. Am. Chem. Soc. 40 1361
Moncton D E and Pindak R 1979 Phys. Rev. Lett. 43 701
Moncton D E and Pindak R 1980 in Ordering in two dimensions (ed.) S K Sinha (Amsterdam: North Holland) p 83
Mor R H 1979 Phys. Rev. Lett. 43 931
Nelson D R 1983b Physica A118 319
Rahman A 1964 A136 405
Rajaraman R 1982 in Solitons and instantons (Amsterdam: North Holland)
Ramakrishnan T V 1982 Phys. Rev. Lett. 48 541
Ramakrishnan T V 1984 Pramana 22 365
Saito Y 1982a Phys. Rev. Lett. 48 1114
Sinha S K 1980 (ed) Ordering in two dimensions (Amsterdam: North Holland). This conference proceedings is representative of the vigorous activity in the field at that time.
Stanley H E 1971 Introduction to phase transitions and critical phenomena (Oxford: Clarendon)
Young A P 1979 Phys. Rev. B19 1855
Young A P 1980 in Ordering in strongly fluctuating condensed matter systems (ed) T Riste (New York: Plenum) p 193
Zollweg J A 1980 in Ordering in two dimensions (ed.) S K Sinha (Amsterdam: North Holland) p. 331

Note added in proof

After this paper was sent for publication, we became aware of:

(ii) a paper by A Holz 1984 (Phys. Lett. 105A 472) who comments on Kleinert’s work, and

Attention is also invited to the Conference Proceedings entitled Melting, Localization and Chaos (editors R Kalia and P Vashista 1982 Amsterdam: North Holland) which contains many papers relevant to the present article.
Defect structure studies using positron annihilation spectroscopy

R KRISHNAN and D D UPADHYAYA
Metallurgy Division, Bhabha Atomic Research Centre, Trombay, Bombay 400085, India

Abstract. The positron annihilation method is a new addition to the range of sensitive complementary nuclear techniques available for materials' research. The preferential sensitivity of positrons towards micro-defect domains which are not assessable by other techniques makes it an attractive tool for many materials science problems. The present paper is intended as a brief introduction on the principle of measurements and its potential is exemplified with the help of results on some metallic and ceramic systems.

Keywords. Defect structure studies; positron annihilation spectroscopy; sensitive complementary nuclear techniques.

PACS No. 61.70; 78.70

1. Introduction

The existence of the positron, the antiparticle of the electron was predicted by Dirac (1930) and experimentally verified by Anderson (1932). An electron-positron pair is unstable and the annihilation photons constitute the signal through which the information on the behaviour of positrons in condensed matter; including their interaction with imperfections, is obtained.

Positron annihilation spectroscopy has many advantages in the study of condensed matter. It is a non-destructive method. It is now well established as a technique for the study and characterisation of defect microstructures (Doyama and Hasiguti 1973; Seeger 1976; Byrne 1980; Wiffen and Spitznagel 1982).

The physical basis of positron interaction in condensed matter is well described by various authors (West 1973; Hautojarvi 1979; Brandt and Dupasquier 1983). A brief account of the positron annihilation mechanism and the principle of measurement techniques is presented here along with some applications in metals and ceramics.

2. Positron dynamics in solids

Positrons are usually obtained from radio-nuclide sources such as $^{22}$Na, $^{64}$Cu, etc. which emit positrons with sufficient energy to penetrate most materials to a few hundred microns. On implantation, positrons get thermalised rapidly ($\approx 1$ psec) by a succession of ionization and phonon scattering events. The thermalisation times are usually much shorter as compared to the mean lifetime of positrons in solids (Kubica and Stewart 1975).
3. Principle of PAS measurements

Positron, either free or bound, interacts with more than one electron in different energy states in a many particle system. Basically two kinds of information are obtained: (1) that
Defect structure studies

related to the electron density and (2) that of the distribution of electrons in momentum space. Thus, three different types of measurements can be performed: (i) positron lifetime or more specifically, the positron decay spectrum; (ii) angular correlation of the direction of the photon emitted during 2γ-annihilation; (iii) doppler broadening of 511 keV annihilation radiation.

The time delay between the injection of positrons (marked by the 1.28 MeV nuclear γ-ray using a 22Na source) and the detection of 511 keV annihilation photon in the solid, measured over a large number (∼ 10^5) of individual events constitutes the lifetime spectrum (Smedskjaer and Fluss 1983). From such time distributions, one can deduce the decay rate of positrons from experimentally distinguishable states in the material. For ionic solids, like alkali halides and metal oxides, generally two or three lifetime components are obtained. The fastest component τ_1 is assigned to free and p–Ps decay, while the longer ones τ_2 and τ_3, to the pick-off annihilation of o–Ps in the crystalline and disordered regions respectively (Bertolaccini et al 1971; Brandt and Paulin 1968).

If the centre of gravity of the annihilating electron-positron pair is at rest, two γ-rays must be emitted with exactly equal energy (511 keV) and in opposite directions. However, because of the non-zero momentum of the pair, the photons deviate from collinearity, i.e; the two γ-rays are emitted in directions π ± θ radians. In addition, the energy of the annihilating photon undergoes a Doppler shift (511 ± ΔE) keV. Since the momentum of the thermalised positron is very small relative to that of the electron, the distribution of θ in angular correlation and ΔE in the Doppler broadening experiments therefore represent essentially those of the crystal electrons, both core and valence. The higher momentum core electrons contribute proportionally more to the largest values of θ and ΔE. Therefore, the γ–γ angular correlation curves and Doppler broadened spectra are both more sharply peaked for solids containing defects and trap positrons, since there are fewer core electrons at the defect site (Doyama 1979).

For quantitative analysis of defects in solids, parameters deduced from PαS measurements are employed with the help of trapping models (Brandt 1967; Bergersen & Stott 1969).

4. PαS study of vacancy concentrations in metals

PαS has a distinct advantage in the study of vacancy concentration in metals, because it is possible to carry out these studies under thermal equilibrium conditions and at very low vacancy concentration (∼ 10^-6) where only monovacancies dominate the whole equilibrium ensemble.

The common observation in metals is that the positron annihilation parameter shows a typical S-shaped curve: a thermal expansion effect at low temperatures which gets superimposed with a vacancy trapping effect at higher temperatures and a saturation near the melting point, where it is assumed that all the positrons get annihilated from the trapping sites (MacKenzie et al 1967; McKee et al 1972; Triftshauser 1975; Seegers et al 1978). Figure 2 shows the typical lineshape parameter dependence for Cu as a function of temperature (Fluss et al 1980). For evaluating E_v, the formation energy of a vacancy, the experimentally measurable quantities F_b and F_v which represent the specific values of positron annihilation parameter, F (eg lifetime,
ratio of the peak height to the area of angular correlation curve, line-shape parameter etc.), in the free and trapped states respectively, are used in an Arrhenius relationship.

$$\ln \left( \frac{F - F_0}{F_0 - F} \right) = -\frac{E_v}{kT} + \ln (A\tau_b)$$

(1)

where $k$ is the Boltzmann constant, $T$ is the absolute temperature, $A$ is a constant and $\tau_b$ is the lifetime for the perfect lattice. The values of $E_v$ determined from PAS technique are found to be in good agreement with those obtained from other experiments (Siegel 1978; Doyama 1979; West 1979).

5. Radiation damage studies

Study of radiation damage is important from the point of view of understanding the behaviour of point defects in solids. The sensitivity of the positron annihilation spectroscopy technique to vacancies and their aggregates has made it particularly useful in elucidating the recovery processes in metals and alloys after neutron and electron irradiation (Mogensen et al 1972; Cotterill et al 1972). From the PAS data, it is possible to identify the migrating entity and determine the average dimensions of vacancy clusters or microvoids, which are below the limit of conventional electron microscopy (Seeger 1973; Petersen et al 1975; Mantl and Triftshauser 1975).

Irradiation at low temperatures predominantly results in the production of single vacancies and interstitials. The defect recovery in copper crystals (Mantl and Triftshauser 1975) irradiated by 3 MeV electrons at 10 K, was followed by residual resistivity and Doppler broadening parameters expressed in terms of $I_v$ (corresponding to valence electrons) and $I_c$ (corresponding to core electrons) (figure 3). The initial drop in resistivity indicates that the number of trapping sites for positrons is reduced. Thus, the increase in $I_v$ and the corresponding decrease in $I_c$ can be interpreted as due to the aggregation of vacancies into small three-dimensional clusters, which eventually collapse into vacancy loops, often observed in electron microscopy studies. These results thus support the vacancy model of stage III for migration of point defects.
6. Defects in non-metals

In non-metals, the positron annihilation modes are quite complex—usually characterised by two or more decay components—and the reason for this has still not been established. However, positron annihilation behaviour in molecular solids like polymers, ionic solids like alkali halides, oxides and the silicate glasses is explained, partly because of the observations made with other conventional techniques (Walker 1979; Dupasquier 1979). As positron trapping by defects is the basic phenomenon, the established analytical methods are equally effective in analysing the PAS results.

The formation of a Ps-like bound state is no longer speculative and several confirmatory experiments, involving measurements under different atmospheres (Steldt and Varlashkin 1972), three-to-two photon relative yield measurements (Sen and Patro 1969), magnetic quenching experiments (Yam et al 1978; Judd et al 1979) etc., have helped to reveal the annihilation mechanism. Figure 4 shows the $\gamma-\gamma$ angular correlation results on pure and doped polycrystalline samples of ThO$_2$. An increase in the peak counting rate in the presence of transition metal impurities represents the enhancement in the 2$\gamma$ annihilation events compared to that in pure ThO$_2$. This is attributed to the increase in the pick-off quenching rate of o-Ps in the presence of these ions.

7. Positron annihilation in fine powders

Positronium formation in metal oxide powders was first reported for Al$_2$O$_3$, MgO, CaO and ZnO (Kusmiss and Stewart 1967). The angular correlation curves showed a low intensity narrow component attributed to the annihilation of p–Ps, formed on or near the surface of the finely divided powder particles. Earlier observations of long time components ($\tau_3 \approx 3.5$ ns, $I_3 \approx 6\%$) observed for alkaline-earth oxides (Bussolati and Zappa 1964) and the additional intermediate component ($\tau_2 \approx 0.5$ ns, $I_2 \approx 12\%$)
obtained in several oxides and fluorides (Paulin and Ambrosino 1967) also gave a clear evidence for the decay of a Ps-like bound state in the medium.

The most remarkable feature observed is that the annihilation characteristics are strongly dependent on the particle size of the powders. For ZrO$_2$ powders (Mitsuhashi et al 1972), it was found that the positron life time is insensitive to its allotropic modification but is dependent only on the particle size. Life time components $\tau_1$ and $\tau_2$ were found to increase as the particle size becomes small.

Brandt and Paulin (1968) demonstrated a strong dependence of life time spectra on particle size of MgO, Al$_2$O$_3$ and SiO$_2$. The systematic narrowing of the angular correlation curve (Steldt and Varlashkin 1972) and the increase in the pick-off life time component $\tau_2$, (Baranowski et al 1977) with decreasing particle size of MgO powders are also consistent with the results of previous authors. The decrease in mean life time with improvement in crystalline perfection is a general phenomenon observed in metal powders (Noguchi and Miyata 1979), graphite (Iwata et al 1981), diamond (Mokrushin and Breusov 1980), etc.

Figure 5 shows the life-time results on microcrystalline ThO$_2$ powders (Upadhyaya et al 1982a). A linear plot correlating $\tau_1$ and $\tau_2$ with crystallite size deduced from X-ray line broadening, established that the finer particles correspond to a higher defect density. The annihilation probability for both free ($\tau_1$) and $\sigma$–Ps pick-off ($\tau_2$) processes is reduced in the pseudo-amorphous powders due to lower than average electron density. With increasing crystalline perfection, the free volume at the trapping site decreases which enhances the overlap of the positron wave function with that of the valence electrons in the medium. Deviation in the value of the positron parameter from that of the single crystal state is thus a quantitative measure of the degree of disorder in the powdered state. This leads to the suggestion that a reexamination of the intrinsic
values of positron life times reported so far by employing powder samples may be needed.

Another important aspect of positron dynamics in fine powders is concerning their diffusion mechanism—both from the point of view of moderation as well as trapping (Barko et al 1979; Brandt 1974). The injected positrons after thermalization and Ps formation scan the powder grains isotropically. The rate of trapping at defects and the escape rate at the surface are proportional to the diffusion constant of the migrating species. Assuming the Ps to be formed with equal probability inside a small insulating solid of spherical shape, the escape rate can be derived. The relative intensity of the life time component assigned to the surface, when expressed as a function of particle size enables one to calculate the Ps diffusion constant \( D_s \). For oxides, it was found to be of the order of \( 10^{-5} \text{ cm}^2\text{ sec}^{-1} \) (Brandt and Paulin 1968; Upadhyaya et al 1982b), which is about three orders of magnitude smaller than for metals (Paulin et al 1974; Matsuoka et al 1980; Noguchi and Miyata 1979).

Positron diffusion constant in powders can also be measured using the following expression (Brandt 1974)

\[
Y = (D_+ \tau_b)^{1/2} S_d
\]

where \( Y \) is a comprehensive life time parameter, \( \tau_b \) the intrinsic value of life time in the solid and the quantity \( S_d \), which is the product of the specific surface area of the powder and the theoretical density, is equivalent to the inverse of the mean particle diameter.

The values of the escape probability \( I_S \) and \( Y \) for thoria powders are plotted in figure 6 (Upadhyaya et al 1982c). It shows an increase in the intensity of the surface life time component with increasing surface area. The slope of the \( Y vs S_d \) graph yields the Ps diffusion constant in thoria as \( 80 \times 10^{-5} \text{ cm}^2\text{ sec}^{-1} \), equivalent to a Ps diffusion length of about 60 A.
Summary

Positron annihilation spectroscopy has made a significant impact in our understanding of the defect structure in solids. The favourable combination of properties possessed by positrons makes them especially valuable as probes of microscopic and atomic scale imperfections. It has been found to be a sensitive tool for the study of electron structure in metals and also has a great potential in the study of point defects—especially lattice vacancies. This selective nature of the interaction has helped in explaining various metallurgical phenomena. The defect recovery processes associated with the irradiation damage and various thermomechanical treatments in metals and alloys have been investigated to great advantage.

PAS has successfully utilized oxide powder systems, primarily as a suitable medium to test and rationalize experimentally the predictions of quantum electrodynamics about the positronium formation and its decay reactions. In this process, it has contributed to our understanding of defect structure in ceramics, in much the same way as it has in metals. Results cited here elucidate the role of material fineness on Ps-dynamics, ie, the diffusion constant, pick-off annihilation rate, etc., employing a medium which has been well characterized by a variety of conventional techniques.

References

Anderson C D 1932 Science 76 238
Bergersen B and Stott M J 1969 Solid State Commun. 7 1203
Brandt W 1974 Appl. Phys. 5 1
Brandt W and Dupasquier A (Eds) 1983 Positron Solid-State Physics (Amsterdam: North-Holland)
Byrne J G 1980 Scr. Metall. 14 3
Defect structure studies


Dirac P 1930 *Proc. Phys. Soc. (London)* **A126** 360


Doyama M and Hasiguti R R 1973 *Cryst. Lattice Defects* **4** 139

Dupasquier A 1979 in *Positron in solids* (ed) P Hautojarvi (Berlin: Springer-Verlag) p. 197


Hautojarvi P (ed.) 1979 *Positron in solids* (Berlin: Springer-Verlag)

Hinode K, Tanigawa S and Doyama M 1978 *J. Nucl. Mater.* **69 & 70** 678


Kubica P and Stewart A T 1975 *Phys. Rev. Lett.* **34** 852


Mantl S and Triftshauser W 1975 *Phys. Rev. Lett.* **34** 1554


Paulin R, Ripon R and Brandt W 1974 *Appl. Phys.* **4** 343


Sen P and Patro A P 1969 *Nuovo Cimento Soc. Ital. Fis.* **B64** 324

Siegel R W 1978 *J. Nucl. Mater.* **69 & 70** 117


Steldt F R and Varlashkin 1972 *Phys. Rev. B5* 4265

Triftshauser W 1975 *Phys. Rev. B12* 4634


Upadhyaya D D, Muraleedharan R V and Sharma B D 1982c *Proc. second national symposium on positron annihilation (Madras)* (in Print)


West R N 1979 in *Positron in solids* (ed) P Hautojarvi (Berlin: Springer-Verlag) p. 89

Wiffen F W and Spitznagel J A (Eds) 1982 *Advanced techniques for characterizing microstructures* (Pennsylvania: Metallurgical Society of AIME)

Radiation-induced structural changes in alloys

K KRISHAN and C V SUNDARAM
Reactor Research Centre, Kalpakkam 603 102, India

Abstract. Development of alloys for reliable performance in extreme radiation environments is vital for the viability of advanced nuclear reactor systems. Over the past decade, there has been a considerable growth in our understanding of the basic processes of radiation damage, the nature of the induced defects, their interaction and migration, and the influence of these on the mechanical behaviour of metals. This understanding has however come mainly from studies in pure metals and dilute alloys, and there are difficulties when applying these concepts to concentrated alloys, particularly of technological interest. The present article, which attempts to bridge this gap, discusses recent research developments and some of the emerging new concepts as applicable to alloy systems. Interstitialcy transport; percolation effects in defect migration; short range and long range ordering and restructuring of alloys; defects and damage behaviour of metallic glasses; synergetic processes and phase instabilities; and finally, swelling, irradiation creep and ductility behaviour of alloy systems are the topics discussed.

Keywords. Alloy systems; interstitialcy transport; percolation effects; defect migration; short range ordering; long range ordering; metallic glasses; swelling; irradiation creep; ductility.

PACS No. 61.80

1. Introduction

Growth of commercial nuclear reactors and development of advanced reactor systems have led to a demand for new alloys which can reliably be used in radiation environments. Considerable advance has been made in our understanding of radiation effects in materials over the past decade for safe and economic operation of nuclear reactors. For the pressurized heavy water reactors (PHWRs) for example, extensive investigations have been necessary to understand the in-reactor behaviour of zirconium alloys, and to improve the performance of zirconium alloy fuel tube and pressure tubes through appropriate control of texture and microstructure during tube fabrication (Sinha and Kakodkar 1983). In the case of fast breeder reactors (FBRs), radiation damage effects in structural components are orders of magnitude more severe because of the increased damage rates due to fast neutrons and embrittlement problems associated with the production of helium by \( (n, \alpha) \) reactions. With the alloy development and evaluation efforts over the last decade (see the reviews by Rodriguez et al 1984; Sundaram 1984) alloys like modified stainless steels, ferritic steels and nimonic alloys have emerged showing reduced swelling, irradiation creep and helium embrittlement behaviour. In more advanced reactor concepts like fusion systems, the damage to the first wall by direct impingement of light and heavy ions can erode the surface besides affecting the mechanical properties (Gruen 1979). Here, along with the development of new alloys (e.g. V-15\% Cr-5\% Ti) new technologies, such as rapid solidification (to produce metallic glasses), formation of surface coatings (e.g. TiC), have also been investigated to ascertain their superior radiation damage resistance.
In all these developments significant contributions have come from the more basic aspects of radiation damage research. There has been a tremendous growth in our understanding about the mechanism of damage and the behaviour of vacancies; self-interstitials; their configurations and clustering kinetics; precipitation; diffusion of gases like helium; and the effect of stress and temperature on these defects (see the review of Ullmaier and Schilling 1980). This has emerged mainly from studies in pure metals, often on single crystals and by investigating simple defect structures. To some extent all this can be extended to dilute alloys but there are serious difficulties while applying these principles to concentrated alloys. Here, even the basic building blocks like vacancies and self-interstitials become diffuse and their structure and properties are not understood. Even well-established facts in pure metals like the high mobility of self-interstitials cannot be assumed since there is gradually emerging evidence that in some alloys they may be slower than vacancies and become mobile only at higher temperature. Emphasis now in basic research is shifting to radiation damage problems in concentrated alloys. This shift is encouraging since it brings basic research into contact with materials of technological importance which have so far mainly remained in the realm of metallurgical research.

Basic research efforts at understanding the radiation behaviour of alloy systems are still in the early stages. This article discusses some of the new concepts which are emerging in this field and the differences from pure metals are highlighted. In §2 the damage mechanisms are discussed together with defects and their mobility in alloys. Section 3 deals with the secondary effects of damage; segregation and dissolution, order-disorder and damage effects in metallic glasses. The highly non-equilibrium nature of the processes involved introduces interesting possibilities of synergetic effects which could manifest as phase instabilities. These are briefly discussed in §4. Finally, §5 relates these structural changes in alloys with their metallurgical properties and behaviour. Accelerators are increasingly being used for such investigations and some of the recent results are presented.

2. Displacement damage and defects in alloys

The process of atomic displacements and the production of the initial damage structure can be discussed under three headings (i) primary and secondary atomic displacements, (ii) defects and their mobility and (iii) mixing effects and structural changes.

2.1 Displacement process

2.1a Primary displacements: The early stages of damage in alloys are essentially similar to those in pure metals. They involve the creation of the primary-knock on-atoms (PKAS) by the incident irradiation and secondary damage as Frenkel pairs and cascades produced in turn by the PKAS. For an incident monoenergetic radiation (neutrons/ions/electrons) with energy $E$, the PKAS acquire a spectrum of energies due to recoils of energy $T$. This energy spectrum is governed by the nuclear scattering and involves the nuclear differential cross-section $d\sigma_n(E, T)$ which depends on the masses and nuclear charges $(M_i, Z_i e)$ and $(M_p, Z_p e)$ of the incident ($i$) and target atom ($P = A, B, \ldots$, for alloy atoms) respectively. In non-relativistic situations the maximum recoil energy transferred is $T_{\text{max}}, T_{\text{max}, p} = 4M_i M_p E/(M_i + M_p)^2$. For neutrons...
due to hard sphere collision of radius \( R \),
\[
\frac{\text{d} \sigma_p(E,T)}{\text{d}T} = \frac{R^2}{7} \frac{\text{d}T}{\text{max},p}
\]
which is independent of \( T \). In the case of charged particles, Rutherford scattering predominates and
\[
\frac{\text{d} \sigma_p(E,T)}{\text{d}T} = \frac{\pi Z_p^2 Z_p^2 e^4 M_p}{\text{d}T} \frac{\text{d}T}{M_p E T^2}.
\]
(1)

In this case there is a strong dependence of \( \frac{\text{d} \sigma_p(E,T)}{\text{d}T} \) on \( T \) showing that most of the energy loss occurs with small recoil energies at approximately right angles to the incident particle direction.

Consider a binary alloy having elements with well separated masses \( M_A \), \( M_B \) (and charges \( Z_A \), \( Z_B \)). The strong dependence on these variables of \( \frac{\text{d} \sigma_p(E,T)}{\text{d}T} \) and \( T_{\text{max},p} \) results in two well-defined PKA spectrums. In an alloy of NiAl for instance, \( \frac{\text{d} \sigma_{\text{Ni}}}{\text{d} \sigma_{\text{Al}}} \sim 2 \) at all recoil energies and \( \frac{T_{\text{max},\text{Ni}}}{T_{\text{max},\text{Al}}} \sim 0.5 \) as seen from figure 1a. An interesting situation arises in the case of irradiation by energetic electrons (\( \gtrsim 0.5 \text{ MeV} \)) and very low energy light ions (\( \sim \text{keV} \)). Due to the large difference in the maximum recoil energies it could so happen that the entire damage process is confined only to the lighter atom species constituting the alloy. A number of alloys like stainless steel which are technologically important contain a significant amount of light elements like C, Al and Si. These elements could be either dispersed or could be in the form of precipitates generically called \( \gamma' \): like Ni\(_3\)Al, Ni\(_3\)Si, and are responsible for the hardening characteristics of alloys. The process of selective displacements upsets the equilibrium of the thermally formed composition resulting in the restructuring of the alloy.

The subthreshold effects dominate at low energies (except for electrons) where the scattering cross-section is high as seen from figure 1a. However at higher energies the displacement damage by the heavier masses becomes more significant as can be seen from figure 1b. There is however, another feature which becomes important in alloys even if there is no mass difference between the alloying elements. This arises from the dependence of the PKA spectrum on the local atomic concentrations \( \xi_A(r) \) and \( \xi_B(r) \),

**Figure 1.** (a) Recoil energy spectrum of Ni and Al atoms in NiAl alloy with 2 MeV protons (b) Comparative damage rates produced by the Ni and Al PKA spectrums for different proton energies (Panigrahi 1984).
The pka spectrum (ignoring mixing effects) is given by
\[ p_{\text{KA,AB}}(r,T) = \left(\frac{d\sigma_{\text{AB}}}{dT}\right) N \xi_{\text{AB}}(r), \]  
where \( N \) is the total atomic concentration. The spatial variation in \( \xi_A(r) \) and \( \xi_B(r) \) arises because the distribution of \( A \) and \( B \) atoms would depend on whether the alloy has stoichiometric order, short range order (sro), long range order (lro) or is completely disordered. Thus even though the irradiation is completely homogeneous, inhomogeneities in the pka spectrum lead to gradients in the defect structure introducing fluxes of atoms and defects which can restructure the alloy (Krishan and Abromeit 1984).

2.1b Secondary displacement effects: The process by which PKAs produce secondary displacements is via replacement collision sequences where each atom displaces the neighbouring one along the crystallographic directions. A vacancy is generally produced at the beginning of the chain and the replacement sequence propagates 3 to 6 lattice distances (sometimes as long as 20) till the extra atom is displaced to an interstitial position. For a replacement sequence the PKA must transfer an energy between \( T_f \) and \( T_d \) where \( T_f \) is the focussing energy and \( T_d \) the minimum threshold displacement energy to form a stable vacancy-interstitial atom pair. \( T_d \) is typically 10–30 eV and depends on the crystallographic direction. The higher limit \( T_f \) arises because of dechannelling effects and for energy transfers larger than these, cascades are formed with several vacancies (\( \gtrsim 1000 \) for typical large cascades) and the atoms are removed to peripheral interstitial sites by several replacement sequences.

The scenario of separation of vacancies from self-interstitials via replacement sequences to escape spontaneous correlated recombination needs to be considered more carefully in alloys. The complete transfer of energy from the moving to a stationary mass, when the masses are equal, forms the basis on which replacement sequences are formed. In alloys, where different masses are involved, there will be a large dissipation of energy during transfer and this effectively inhibits the propagation of such sequences. Figure 2 shows four different situations which can arise and these will be discussed.

Figure 2a shows an ordered FeAl alloy. In this alloy we find that channels along the \( \langle 111 \rangle \) direction are blocked and no replacement sequences can propagate; on the other hand the \( \langle 100 \rangle \) and \( \langle 110 \rangle \) channels are free. The threshold displacement energy surface will show a high degree of anisotropy in such alloys. There have been very few detailed studies of such effects though some resistivity measurements have been made at low temperatures (Rivier and Dinhut 1981) after irradiating with electrons along various crystal orientations. Figure 2b shows a plane of atoms for a NiAl alloy. Such alloys are composed of two interpenetrating simple cubic sublattices \( \alpha \) and \( \beta \). On the A-rich side the excess A-atoms are accommodated on the \( \beta \)-sublattice as antistructure defects, while on the B rich side structural vacancies are formed on the \( \alpha \)-sublattice. The existence of these vacancies which can be as much as 10% would not make it possible for any collision sequences to operate on the \( \alpha \)-sublattice as shown in the figure 2b. Such vacancies would play a very effective role as recombination centres and therefore significant differences can be expected in the formation of the microstructure as a function of alloy composition. In another class of alloys viz CuZn and AuCo the lattice defects consist of A-atoms on \( \beta \)-sites, B-atoms on \( \alpha \)-sites and vacancies, the behaviour of replacement sequences would be correspondingly different. A more complex
situation arises in cases where the alloy is disordered as shown in figure 2c. In such cases if the mass difference is large it should be very difficult for replacement sequences to operate along any direction making them ideal radiation damage resistant materials. However, this is not the case since some amount of short range ordering will persist. Since it is sufficient for chains to be a minimum of three to four lattice spacings long to form a stable Frenkel defect, a large number of chains would be possible. However, under identical conditions resistivity measurements at low temperatures do indeed show that the disordered phase is more radiation resistant than the ordered one. Finally, figure 2d shows the even more extreme situation encountered in metallic glasses. In this case besides composition disorder, one encounters spatial disorder in the atomic arrangements. Here also there is evidence (Egami 1983) for chemical stoichiometric composition vacancies appearing as structural defects on one of the sublattices blocking all replacement channels in the sublattice. The shaded dots are on a separate plane and belong to a different sublattice. (c) In a disordered alloy with unequal masses all channels would be blocked but for the short range clustering which allows replacement sequences to occur making them also vulnerable to damage.

2.2 Defects and their mobility in concentrated alloys

We have already encountered two defects namely 'structural vacancies' and 'antistructure' defects which exist when we move away from the stoichiometric compositions in
alloys like NiAl, CoGa, CuZn, AuCd. Interstitialcy transport via the dumb-bell type configuration has been identified as a possible mechanism for transport though there is only limited experimental evidence. Vacancy-interstitial complexes also appear to play an important role in some alloys like Fe-Cr-Ni containing impurities. On the other hand, in systems like metallic glasses even less is known.

2.2a Interstitialcy mechanisms: Our understanding of defects in concentrated alloys comes mainly from the experience in dilute alloys and pure metals. Here the self-interstitial is known to exist as a dumb-bell along <100> direction in fcc lattices. In a concentrated alloy A, B this picture would involve three possible dumb-bells A-A, A-B and B-B. If their binding energies and migration energies are about the same they would be mobile and would freely change their form as shown in figure 3. Due to such continuous exchanges the identity of the interstitial is lost though atomic mobility can take place. Identification of such a process experimentally is difficult since the behaviour is very similar to a self-interstitial in a pure metal and produces no additional structure in resistivity measurements. In principle such a complex could lead to segregation though there is no direct experimental evidence for this. Such a mechanism is believed to operate in alloys like Cu Ni (Wollenberger 1982) possibly in stainless steel Fe-Cr-Ni, and alloys of Cu with Co, Fe, Zn and Ge (Robrock 1983).

A more complex situation arises if the binding energies for the A-A, A-B and B-B complexes are relatively different and so are their migration energies. Three possibilities exist as shown in figure 4a depending on \( \omega_t \), the trapping and \( \omega_d \), the detrapping jump frequencies. In this case the ability of the defect to move becomes strongly dependent on the paths available as shown in figure 4b. Bocquet (1983) showed that when the trapping is infinitely deep \( \omega_t \gg \omega_d \), the ability for the defect to percolate depends on the solute concentration. If the solute concentration \( C_B \) is lower than a limiting percolation value \( p_c \approx 0.29 \), the defect is trapped in clusters of B-atoms which are entirely surrounded by A-atoms and the long range migration of the defect becomes impossible. Internal friction experiments, done by Halbwachs and Beretz (1980), show that in AgZn the interstitial mobility is smaller than the mobility of vacancies. Generally strong trapping effects are observed when an oversized "impurity" atom is involved or an undersized 'impurity' is involved. A detailed discussion of this has been given in the review by Robrock (1983).

2.2b Minor alloying addition in Fe-Cr-Ni alloys: Fe-Cr-Ni alloys are technologically important and elements like Ti, Si, Mo have been used to control their swelling behaviour. Sharma et al (1978) and Dimitrov et al (1981) have obtained some

![Figure 3](image)

*Figure 3*. Changes in the dumb-bell configurations as it migrates (a) B-B dumb-bell changing to (b) A-B dumbbell and (c) A-A dumb-bell. Transport by such interstitialcy mechanism also causes structural rearrangements.
Radiation-induced structural changes in alloys

Figure 4. (a) Relative binding and migration energies which can lead to transport by (1) B-B (2) A-A & A-B and (3) A-B dumb-bells $\omega_d$ and $\omega_t$ are the dissociation and trapping frequencies. (b) Percolation of a B-B dumb-bell at low A-atom concentrations and no percolation at high A-atom concentrations. Percolation paths depend on the alloy concentration and effect the atomic mobility by interstitialcy mechanism.

Figure 5. Resistivity change by minor alloy elements indicating different trapping processes (Sharma et al 1978).

information about defect trapping mobilities with these solutes, using low temperature resistivity technique. The main results are shown in figure 5. At low temperatures till 190 K the decrease in the resistivity is due to the recombination of close pair Frenkel defects. Between 190 K and 260 K the transport by the interstitialcy mechanism is postulated in pure Fe-Cr-Ni. However, from the shift in the temperature for Si and Mo, it follows that they strongly trap the interstitials while Ti does not. In the temperature range 260 K to 460 K, vacancies also become mobile and are strongly trapped by Ti and weakly by Si and Mo. At temperatures above 460 K the release of interstitials and vacancies from Si and Ti atoms is responsible for the resistivity decrease. These features are in general agreement with the swelling behaviour in these alloys. It is perhaps worth pointing out that the contribution to resistivity in concentrated alloys comes from the defect concentrations and structural changes.
2.2c Defects in amorphous alloys: There is even less understanding about defects in amorphous alloys more commonly known as metallic glasses. The glassy phase always has a density lower than the crystalline phase indicating that there should be a large free volume available. Experiments have failed to show any presence of large atom sized holes which could be called 'vacancies' and it is therefore thought that this free volume is distributed amongst the atoms. Irradiation would produce instantaneous 'vacancies' due to displacement of atoms but such vacancies exist for only $\sim 10^{-10}$ sec as shown by computer simulation (Bennett et al 1979) and this volume gets distributed (figure 6a). However very low temperature resistivity experiments show some structure and the possibility that they may be stable in some complex form cannot be ruled out. Cahn (1978) on the basis of diffusion data argued that a duality picture involving both distributed volume and atom-sized holes may be required. A theoretical justification for this has been given by Krishan (1982b) on the basis of a random network model where the distributed volume could exist as a configurational lock which could be thermally activated to form an atom-sized hole (figures 6b and c). Such a 'hole' would migrate but would not be detectable since it would collapse into a configuration lock in $\sim 10^{-10}$ sec. The presence of such 'holes' is required to explain atomic transport which appears to have the same temperature dependence as in crystalline metals where the vacancy mechanism is involved. On the other hand, there is not so much difficulty in identifying interstitials since they would reside in the interatomic spaces of the disordered network (shown in figures 6b and c). They cannot form a part of the network since such a process would make the local atomic density to become more than that of the crystalline solid, and there seems to be no evidence for this. Though the structure and properties of interstitials are not understood, the diffusion constant for light gases like He seems to be very similar to that in crystalline solids.

There is no evidence at all for any extended defects and electron microscopy does not reveal any structures which could be equivalent to dislocations, precipitated phases or grain boundaries. Theoretically the possibility of disclinations as stable defects has been postulated by Rivier (1979) but there is no experimental evidence for this. However slip bands have been observed in mechanically deformed metallic glasses.

2.3 Atomic mixing and restructuring in alloys

Besides displacement of atoms, irradiation also rearranges the atoms during replacement sequences since atoms get shifted in the process. This rate is often even greater

![Figure 6](https://example.com/fig6.png)

*Figure 6.* Schematic diagram showing (a) lower density of the glassy phase (b) The same atoms are shown on the basis of an interpenetrating network model with a displacement event a 'vacant site' and an interstitial (arrow) in the internetwork position (c) The vacant site collapses in $\sim 10^{-10}$ sec forming a configuration lock distributing the vacant volume. Structure in (c) by thermal activation can produce the structure in (b) showing the duality behaviour of the 'atom sized hole' and distributed volume (Krishan 1982b).
Radiation-induced structural changes in alloys

than the displacement rate; the ratio \( \varepsilon \) being about 2 for electron irradiation and 80 for fast neutron irradiation in Ni\(_3\)Mn. Also such rearrangements can be caused by cascades since the atoms during the collapse of the cascade will be rearranged. The extent of rearrangement or disorder depends on the state of the alloy. In case there is only short range clustering, then the changes can be quantified in terms of the Warren-Cowley short range ordering (sro) parameters \( \alpha_i \). These parameters express the number of like atoms which on an average would be in \( i = 1, 2 \ldots \) neighbouring shells. On the other hand if the original state is ordered with atoms prescribed to fixed \( \alpha, \beta \) sublattices in the ideal state, then the degree of order is quantified by the Bragg-William lro parameter \( S \) defined by

\[
S = (f_A - X_A)/(1 - X_A),
\]

(3)

\( f_A \) is the probability of finding an \( A \)-atom on an \( \alpha \)-sublattice and \( X_A \) is the average \( A \)-atom concentration. From purely statistical considerations the disordering rate for lro is

\[
(dS/dt)_{irr} = -\varepsilon KS,
\]

(4)

where \( K \) is displacement rate. The changes in sro and lro by irradiation are studied by diffuse and small angle x-ray and neutron scattering techniques. Such measurements reveal the basic mechanisms operating in the alloy. The irradiation effect can be seen in terms of the average distance an interstitial atom is moved from its vacancy in a replacement sequence. This can be expressed by \( W(r - r') \) which is the normalized probability for producing \( A \) (or \( B \)) interstitial at a distance \( (r - r') \). The spatially dependent dose rate \( K_A(r), K_B(r) \) for interstitial production is given by

\[
K_{A,B}(r) = \alpha \sigma_{A,B} \int W(r - r') \tilde{\xi}_{A,B}(r') d^3r',
\]

(5)

where \( \sigma_{A,B} \) is the displacement cross-section for \( A, B \) atoms and \( \alpha \) the proportionality constant. One can get an idea about the restructuring of the alloy if we take the Fourier transform \( W(k) \) of \( W(r - r') \) and decompose it into correlation effects of different wavelengths. This gives some idea about the state of lro which will be most affected since these would have a wavelength comparable to the dominant wavelength in \( W(k) \) (Abromeit and Krishan 1984).

3. Structural changes in alloys

Alloys under irradiation are essentially non-equilibrium systems. They develop concentration gradients in point defects and complexes formed by irradiation. These gradients cause fluxes of atoms and violate the thermodynamic requirement that, at equilibrium, the chemical potential of each species is constant throughout the system. Thus these fluxes produce segregation and restructuring of the alloy resulting in development of new phases, changes in the short range and long range order and new types of sinks to maintain a steady state. In this section we give examples of these effects.

3.1 Process of segregation and precipitation

Radiation-induced segregation has been extensively studied because of its important technological implications. For example, austenitic grains of stainless steel can transform entirely into ferritic during neutron irradiation (Potter 1983), because of Ni
segregation towards and chromium segregation away from grain boundaries. Also despite small concentration of silicon in stainless steels, it tends to segregate to grain boundaries to form intermetallic silicide phases like γ'-Ni₃Si which have a profound influence on the in service performance of components. Segregation essentially occurs via two processes. The inverse Kirkendall effect where a gradient in vacancy concentration near a sink results in a reverse flux of A and B-atoms, the atom diffusing faster leaves behind a region rich in the slower atomic species. The second process involves the diffusion via the defect-solute complex. The complex diffuses in its own concentration gradient which could be opposite of that of the solute. A criteria for such a process is that

$$E_b^\text{complex} + E_m^\text{defect} > E_m^\text{complex}$$  \hspace{1cm} (6)

where $E_b$ and $E_m$ refer to the binding and migration energies of the defects involved.

Converse processes like the dissolution of existing precipitates are also known to occur. This has been observed in Ti-modified stainless steel containing γ' precipitates by Lee et al (1980) after neutron irradiation. The electron micrograph in figure 7 (Varatharajan et al 1982) shows another example of dissolution of γ' precipitates Ni₃(Al, Ti) after 50 keV helium ion irradiation where the fragmentation effects are clearly evident. There are various processes by which dissolution can occur. For example a phase with a positive volumetric mismatch with respect to the matrix will tend to have a bias for vacancies and hence will establish a gradient of those atomic species which migrate by vacancy-mechanism. Another mechanism of dissolution has been postulated by Nelson et al (1972) where the solute atoms are ejected from the precipitates by impinging energetic pka's. The recoil-induced concentration of solutes may exceed the equilibrium concentration and thus would start a new nucleation for the precipitate. Thus it appears that the original precipitate is fragmented. The different stages of this process are schematically shown in figure 8.

3.2 Short-range clustering and long-range decomposition of Ni-Cu alloy

The Ni-Cu alloy has been extensively studied and shows complete miscibility in the solid state at temperatures well above 600 K where thermal diffusion is high for mass transport. At around 600 K thermally-activated mass transport becomes sluggish and

![Figure 7. Dissolution and fragmentation of γ' precipitates Ni₃(Al, Ti) produced by low energy α particles. (a) unirradiated (b) irradiated in nimonc 90 alloy (Varatharajan et al 1982).]
Radiation-induced structural changes in alloys

Figure 8. Schematic representation of fragmentation and reprecipitation of precipitates at different stages.

has a tendency to show short-range clustering (Vrijen 1977). On the basis of this spinodal decomposition (see the review by de Fontaine 1979) is predicted but has never been observed because of low atomic mobility. Wagner et al (1980, 1982) have shown by small-angle neutron scattering experiments on $^{62}$Ni-41% $^{64}$Cu alloy that in the temperature range 373–480 K a long-range periodic composition fluctuation can be produced by electron irradiation, with a wavelength of about 45 Å. Two points of view can be adopted. The role of the irradiation is only to enhance diffusion but the kinetics is determined by the Cahn (1966) and Cook (1970) theory of spinodal decomposition. In this theory the increase in the intensity of the fluctuation of wavevector $k$ is governed by the ‘amplification factor’ $a_{\text{therm}}(k^2)$. This is positive inside the spinodal due to a change in sign of the second derivative of the free energy. Such a model however is not completely consistent with the experimental results (Wagner et al 1982).

The second point of view is to consider the decomposition as a radiation-induced process. This possibility has been examined by Krishan and Abromeit (1984) and Abromeit and Krishan (1984). In the presence of irradiation there is radiation-enhanced vacancy and interstitial diffusion. It would appear that the vacancy mechanism will contribute to an enhancement of the thermodynamic decomposition while the interstitial diffusion would effect the irradiation-induced processes. In the simplest case where point defect annihilation is only by recombination, the vacancy concentration would be the same order as that for interstitials. However due to the much higher diffusion constant of the latter, the interstitialcy transport mechanism will dominate. The thermodynamic amplification factor will now be replaced by a new effective amplification factor.

$$a_{\text{eff}}(k^2) = a_{\text{therm}}(k^2) + a_{\text{irr}}(k^2).$$

The contribution to $a_{\text{irr}}(k^2)$ comes from two effects which have already been discussed, viz, the spatial dependence of $\xi_{A,B}$ which produces fluxes of $A$ and $B$ atoms contributing to restructuring of the alloy and the mixing effect discussed in §3.3 which influence fluctuations of wavevector $k$ depending on $W(k)$.

3.3 Order-disorder transformations in Ni$_4$Mo

The alloy Ni$_4$Mo shows a wide range of order-disorder transformations and is therefore a good example of such processes. It has been extensively studied by Banerjee et al (1984) over a wide range of temperatures 50 K $\leq T \leq$ 1050 K in a high voltage.
electron microscope. During irradiation we have already discussed some of the replacement processes by which the alloy disorders and loses its short range and long range order. The process of irradiation is non-equilibrium in nature and eventually there are competing ordering mechanisms, and the alloy attains some equilibrium structure depending on temperature. Interestingly, the final state to which an alloy develops seems to be very sensitive to the initial state of the alloy as shown by figure 9 where the transition from the long range ordering stage to short range ordering is forbidden. An ordering mechanism map which has been developed in terms of the long range ordering parameters and the experimentally determined short range ordering reflection intensity ratio $R$ is shown in figure 10, which gives the broad features of the effects in this alloy.

A detailed understanding of all the processes responsible for these order-disorder effects is not available. However, the formation of the different regions appear to be related to the different processes by which atomic mobility can occur. At temperatures

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure9.png}
\caption{Allowed and forbidden transitions in order-disorder transformations.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure10.png}
\caption{Radiation ordering Map in Ni$_4$Mo by electron irradiation. S is the long range order parameter and R a measure of short range order (Banerjee et al 1984).}
\end{figure}

Region A = destruction of LRO.
Region B = no significant change if initial state corresponds to $S = 1$ or $S = 0$.
Region C = continuous ordering by decay of the SRO waves and simultaneous amplification of the LRO waves.
Region D = initially as in Region C. After SRO is destroyed completely the degree of LRO increases further by domain formation and growth.
Region E = nucleation and growth of $D_{1a}$ domains.
Region F = destruction of SRO.
Region G = reduction of the degree of SRO and at $T > 550$K transition to LRO.
Region H = development of SRO. Data for the build-up of SRO in the completely disordered material is so far only available for $200$K $< T < 450$K.
Region I = destruction of SRO and transition to LRO.
Radiation-induced structural changes in alloys

At temperatures below 200 K, radiation-induced diffusion is the primary mechanism for atomic mobility. Above this temperature, the alloy ordering map in Figure 10 shows two critical temperatures, $T^*_{(sro)} \approx 200$ K and $T^*_{(lro)} \approx 450$ K. It appears that above $T^*_{(sro)}$, short-range ordering processes are restored in the alloy due to diffusion via an interstitialcy mechanism involving Mo atoms. This is conjectured from the fact that the Mo atom is considerably larger than the Ni atom, which is known to be mobile at somewhat lower temperatures. The temperature $T^*_{(lro)}$ is connected with mobility of atoms via the vacancy mechanism. Banerjee and Urban (1984) have done a detailed calculation based on this process where they have considered the various jump mechanisms from the $\beta-\beta$, $\beta-\alpha$, and $\alpha-\alpha$ sub-lattices which are occupied by the Ni and Mo atoms. These jumps lead to ordering of the alloy and provide an explanation of the high temperature behaviour in ordering mechanism map.

3.4 Irradiation effects in metallic glasses

Metallic glasses are a rather extreme example of alloys where the structure is both spatially and compositionally random. Though a large number of alloys have been produced in the glassy phase, metal-metal glasses like Ni$_{60}$Nb$_{40}$ and Ni$_{64}$Zr$_{36}$ are better suited for radiation environments because of the low neutron capture cross-sections and high crystallization temperatures. Also, they are expected to show high radiation damage resistance because practically no restructuring effects would influence these alloys!

These expectations have to some extent been borne out by experiments. For example, there has been no direct observation of void formation for which the existence of biased sinks like 'dislocations' are required. Rechtin et al. (1978) have studied Ni$_{60}$Nb$_{40}$ using 3 MeV Ni$^+$ ions and have reported a high stability of the amorphous phase and no observation of voids using TEM. Also, Cahn et al. (1981) reported on neutron irradiation of Ni-Zr glasses, they found Ni$_{64}$Zr$_{36}$ to be highly stable whereas alloys with higher Zr concentrations (Ni$_{33}$Zr$_{67}$) partially crystallized. Detailed studies have been made of the partial atomic structure factors of metallic glasses after irradiation using x-ray scattering. Also, other techniques like positron annihilation and resistivity have been used but nothing specific could be concluded regarding the damage structure (see the review by Nandedkar and Tyagi 1984).

The situation under low energy irradiation using gaseous ions is somewhat different. These irradiation conditions broadly fall in the domain of fusion reactors. Here, one finds that though metallic glasses are more resistant to damage effects than crystalline alloys, their behaviour shows a similar pattern (Tyagi et al. 1983a, b, 1984). Gases like helium, hydrogen, and argon have been found to precipitate into gas bubbles which ultimately leads to blistering and flaking. Figure 11a shows the precipitation of helium into bubbles 20–50 Å in size at a very high density of $\sim 10^{19}$ cm$^{-3}$ in Ni$_{60}$Nb$_{40}$ glass. Figure 11b shows blistering in Fe$_{40}$Ni$_{40}$P$_{14}$B$_{6}$ and figure 11c shows flaking in Ni$_{60}$Nb$_{40}$ glasses. Measurements of the critical dose for blistering show that it is 50–100% higher in the glassy phase than in the corresponding crystallized phase.

The reasons for this increased resistance of metallic glasses are not very clear and why the damage structure should eventually be so similar to crystalline materials during gaseous ion bombardment only is also rather surprising in some respects. The higher resistance of metallic glasses to gaseous ion implantation effects appears to be related to the lower ability of such glassy materials to retain insoluble gases like helium as
Figure 11. Electron micrographs showing gaseous implantation damage in metallic glasses:
(a) Helium bubbles (Ni$_6$,Ni$_{49}$,Pt$_{18}$,P$_{12}$), (b) Blisters (Fe$_{40}$,Ni$_{17}$,P$_{18}$,B$_{12}$), (c) Erosion (Ni$_6$,Nb$_{0}$,Ti$_{30}$), (d) $\alpha$-Ti (1983, 1984).
Radiation-induced structural changes in alloys

compared to crystalline materials where they can be easily trapped in vacancies. Though metallic glasses have 1–2% more free volume (equivalent of \( ~10^{20} - 10^{21} \) vacancies!) than the crystalline phase, this as we have discussed, is distributed over several atoms and is not effective in trapping helium which seeks 'atom-sized holes'. However since gaseous ion implantation is accompanied by the production of 'vacant sites' and 'displaced atoms' the helium could be trapped into some of these 'vacant sites' before they collapse in \( ~10^{10} \) sec. Once of course helium is trapped into such a site in a dynamic process, it would stabilize the 'vacancy'. Such a kinetic trapping model has been examined and the main effect of the vacancy collapse is to reduce the rate of 'helium-vacancy' complex formation (Krishan et al 1984a). This explains the higher critical doses for blistering which are observed. Our understanding of the basic damage processes in these glasses is slowly improving.

4. Phase instability by irradiation

For a variety of reasons materials in radiation environments must be looked upon as non-equilibrium systems even though they may be in a steady state condition. This is because they are open systems, continuously subjected to a flux of energetic particles across boundaries, with energies often \( 10^7 \) times more than the thermal energies of the atoms in the solid. Such a system clearly cannot be viewed as a thermodynamic system whose final equilibrium state is governed by minimum entropy or energy considerations but not the path it takes to reach this state. Also the processes involved in irradiation are basically irreversible since irradiation produces point defects which are 'dissipated' at sinks with no reverse fluxes. Thus the basic principle of detailed balance, where microscopic processes must proceed at the same rate both ways, is violated.

4.1 Mechanism of instability

Application of non-equilibrium methods to the behaviour of metals under irradiation is of relatively recent origin. Non-equilibrium systems maintained at steady-state can undergo non-equilibrium transitions when some control parameter crosses a certain critical value. The view that all irradiation effects like segregation, precipitation and ordering are simple consequences of growth or shrinkage by point defect fluxes cannot be maintained, since the highly nonlinear nature of the processes could drive the system to new nonequilibrium states.

As a simple illustrative example consider the growth of 'defects' like precipitate or voids due to a steady flux of 'point defects' which is sustained by the irradiation. In such a situation, a balance will be maintained between the steady-state concentration of 'point-defects' and the size and concentration of 'defects' as it evolves. It is important that this balance should be a stable one \( i.e. \) if we change one of these variables, the system must attempt to come back to its earlier steady state. This can be quantified in terms of the time \( \tau_i \) or frequency \( |\omega_i| = 1/\tau_i \), it takes for it to return to equilibrium, \( i \) being the variable which is disturbed, or the degree of freedom of the system. If this disturbance is localized then it could be decomposed into a disturbance involving several wave lengths or wave vectors by Fourier expansion. Each of these wave vectors will have its own response time or frequency \( |\omega_i(k^2)| \). These frequencies of course depend on the basic system parameters like the sink strengths, recombination constants, dose rate, mixing effect etc and this can be explicitly shown as a dependence on the system variable \( s \) by
\(|\omega_{l}(s, k^{2})|\). Near the steady state the approach to equilibrium is exponential in time \(t\) so that the system will re-establish its equilibrium as
\[
\sim \exp\left[\omega_{l}(s, k^{2})t\right].
\] (7)

An interesting situation can arise if for a given \(k\) value \(\omega_{l}(s, k^{2})\) becomes zero for \(s = s_{c}\), a critical value. This would mean that the time \(\tau_{l}(s, k^{2})\) taken for wavelengths fluctuations \(k\) to return to their original steady state would be infinite. In short the system would never return to its original steady-state but will bifurcate to a new non-equilibrium state and a new structure will evolve. For a stable system all \(\omega_{l}(s, k^{2})\) are negative implying a decay to steady state.

4.2 Physical systems

Several examples of such processes have been considered in recent years. Barbu et al (1981) discussed a precipitation instability in undersaturated Ni-Si alloy. They found that there is a threshold value of dose rate which depends on temperature.

A similar instability has also been reported in AlZn alloy by Cauvin and Martin (1983). Another interesting application of this theory is the void-lattice formation. Voids are initially disordered but after certain time orders into a lattice with the same symmetry as that of the host matrix. Besides pure metals this is formed in a number of alloys including stainless steel. In this case the disorder to order transition occurs at a critical value of the sink densities \((\rho_{s}c)\) and \((\rho_{v}c)\) for voids and dislocations like vacancy loops respectively.

Taking into account the various rate processes it is found that \(\omega_{l}(s_{c}, k^{2}) = 0\) for \(s_{c} = \left[\left(\rho_{s}c\right), \left(\rho_{v}c\right)\right]\) and this is shown in figure 12. The value of the wavelength \(\lambda_{c} = 2\pi/k_{c}\) in this model depends on the mean free path of point defects before they are absorbed.

![Figure 12. Critical sink densities \(\rho_{s}, \rho_{v}\) for void ordering in stainless steel (Krishan 1980, 1982a).](attachment:image)
Radiation-induced structural changes in alloys

The long range ordering under electron irradiation in NiCu alloy which was discussed earlier would involve a similar process since the initial short range clustering eventually gives way to long range ordering which is characteristic of such dissipation structures. In this case the mixing effect of interstitials is identified as the order parameter (Abromeit and Krishan 1984).

5. Changes in properties of alloys

Besides radiation, alloys used in reactors are subjected to high stresses and temperatures. Their stability under all these three influences must be ensured. We have already seen a number of structural changes which alloys undergo during irradiation. These changes influence the mechanical behaviour and generally have a degrading influence. It is therefore not sufficient to understand the changes in the alloy structure by irradiation but also to correlate these with its changes in properties like creep, swelling, ductility and fracture behaviour. In table 1 some of these features have been tabulated. One of the problems in such studies in neutron environments is the difficulty in monitoring under controlled conditions parameters like stress and temperature. Therefore, accelerators are being increasingly used for such metallurgical investigations (Krishan et al 1984b) and we will give a few illustrations of such studies.

In §3 we had seen from resistivity studies, the influence on defect mobility of elements like Si, Mo and Ti (figure 7). Gessel and Rowcliffe (1977) discussed the

<table>
<thead>
<tr>
<th>Phase change</th>
<th>Affected properties</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formation of finely dispersed precipitates (e.g., γ')</td>
<td>Flow properties such as yield strength, ultimate strength and work hardening rate</td>
<td>(a) Precipitation hardening (b) loss of solid solution hardening and (c) changes in the kinetics of recovery and recrystallisation processes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(a) Dislocation pinning (b) removal of 'swelling inhibitors' from solution (c) precipitate collector effect.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formation of 'massive' phase</td>
<td>Flow properties such as yield strength, tensile strength and work hardening rate</td>
<td>(a) Second phase strengthening (b) changes in the kinetics of recovery and recrystallisation processes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(a) Loss of 'swelling inhibitors' from the parent phase.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Formation of semicontinuous or continuous films along grain boundaries or interphase boundaries</td>
<td>Impact toughness, fracture toughness and ductility</td>
<td>Propagation of cracks along paths having low interfacial energy.</td>
</tr>
<tr>
<td>(b) Segregation along grain boundaries or interphase interfaces</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
influence of these and other elements like W, Al, Nb and C on Fe-7.5% Cr-20% Ni alloy. Their results are shown in figure 13a and show a marked reduction in the swelling. This is ascribed to the trapping of vacancies and interstitials by the impurity additions. However, these effects can manifest in a complex way since segregation and precipitation could remove these elements from the matrix. The formation of γ'-Ni₃(Ti, Al) could for instance, remove Ti and Al which would result in high swelling rates at higher doses. This is demonstrated from the work of Bajaj et al (1980) where the peak swelling is shown to increase with the volume fraction of γ' in nimonic PE16 alloy (figure 13b).

The influence of stress on swelling has been a matter of considerable concern. It had so far been difficult to establish this effect and to make a systematic study because of the complex loading patterns encountered in reactors. Such effects have been recently studied by Khera et al (1982) and Sahu and Jung (1984) using light ion irradiation from cyclotrons in the energy range 7–10 MeV. The studies on 'pure stainless steel' are at low doses where the influence on the early stages of void growth can be studied. Under tensile stresses ~ 100 MPa the results show (figure 14) a marked increase in the void density. It is interesting to conjecture the effect of a compressive and a tensile stress on void nucleation. As seen from figure 14 the void size distribution is significantly influenced there, being larger number of voids with larger diameters. This is presumably because under a compressive stress void nucleation is inhibited while a tensile stress aids it.

Similar studies on the irradiation creep behaviour under both compressive and tensile stress have been done by Sahu and Jung (1984) in 'pure stainless steel' and the results for compressive stress are shown in figure 15. The strain rate attains a steady state value after approximately 10 hrs of irradiation. This is presumably because in the early stages of irradiation the alloy restructuring takes place and sinks like voids develop. This region is the primary irradiation creep region.

As a final example we discuss some results on Fe-3.5% Ni ferritic steel (A 203 grades). This steel has been irradiated at the cirus reactor to a neutron fluence of 3.5

![Figure 13](image-url)

Figure 13. (a) Influence of minor alloying elements on swelling. The swelling ratio is normalized with respect to the 'pure' alloy taken as unity (after Gessel and Rowcliffe 1977) (b) Effect of γ' precipitation on swelling due to removal of minor alloying elements (after Bajaj et al 1980).
Radiation-induced structural changes in alloys

Figure 14. Comparison of void size distribution during compressive and tensile stress (Sahu and Jung 1984). Inset shows the effect of dose and tensile stress on void swelling (Khera et al. 1982). Experiments have been done using light ion beams at 10 MeV with accelerators.

Figure 15. Strain due to irradiation creep produced by compressive load using 7 MeV protons (Sahu and Jung 1984).

\[ \times 10^{19} \text{ncm}^{-2}, \] and investigated for its ductility and impact toughness under irradiation. The nil ductility transition temperature NDTT should remain nearly 40°C below the operating temperature of any component. NDTT was measured by impact energy experiments and the results are plotted in figure 16 for the irradiated and unirradiated sample (Baldev Raj et al. 1983). A significant increase in the NDTT is observed from a value of \(-91^\circ \text{C}\) to \(149^\circ \text{C}\). Due to irradiation the flow stress of a

P.

32
material is increased while the fracture stress is reduced. Field ion microscope studies show nucleation of very small loops probably arising out of the interaction of solute atoms with irradiation produced interstitials.

References

Abromeit C and Krishan K 1984 to be published
Baldev Raj, Kasiviswanathan K V and Rodriguez P 1983 Reactor Research Centre, Kalpakam report RRC-60
Bocquet J L 1983 Philos. Mag. 47 547
Cahn J W 1966 Acta Metall. 14 1685
Cahn R W 1978 C R 21 Colloque de Metallurgie Speciale, Saclay, France
Cauvin R and Martin G 1983 J. Nucl. Mater 83 67
Cook H E 1970 Acta Metall. 18 297
Egami T 1983 in Amorphous metallic alloys (ed.) F E Luborsky (London: Butterworth) p. 100
Gaskell P H 1980 Nature (London) 289 474
Halbwachs M and Beretz D 1980 in Third European conference of internal friction and ultrasonic attenuation in solids (Manchester) p 139
Krishan K 1980 Nature (London) 287 420
Radiation-induced structural changes in alloys

Krishan K 1982a Radiat. Eff. 66 121
Krishan K, Tyagi A K and Nandedkar R V 1984 (to be published)
Nandedkar R V and Tyagi A K 1984 in Metallic glasses: production properties and applications (ed.) T R Anantharaman (Switzerland: Trans Tech Publications Ltd.) p 165
Panigrahi B 1984 Private communication
Rivier N 1979 Philos. Mag. 40 859
Sahu H K and Jung P 1984 (to be published)
Sundaram C V 1984 Sixth IAEA RCA Working Group Meeting, Reactor Research Centre, Kalpakkam-603102
Ullmaier H and Schilling W 1980 in Physics of modern materials (Vienna: IAEA publication SMR-46/105) p 301
Varatharajan K, Tyagi A K and Nandedkar R V 1982 in Proc. DAE symposium on nuclear physics and solid state physics 25C 562
Vrijen J 1977 E C N Petten Report ECN-31
X-ray photoelectron spectroscopy for surface film analysis in corrosion research

P K CHAUHAN, H S GADIYAR and R KRISHNAN
Metallurgy Division, Bhabha Atomic Research Centre, Trombay, Bombay 400085, India

Abstract. Surface films on metals and alloys often protect them from reaction with the environment, and hence a knowledge of their protective properties and composition could be invaluable for predicting their corrosion behaviour. XPS (x-ray photoelectron spectroscopy) could provide a quantitative analysis of the chemical composition, the nature of valence states and elemental distribution within the surface films.

The present paper reviews the potential of this technique in corrosion studies. A brief review of the work done on the passivation of iron and iron-chromium alloys and on the inhibition studies on copper base alloys has been given. A few examples of investigations carried out at authors' laboratory are also included. An attempt has been made to establish a correlation between the compositions of the films formed and corrosion behaviour of carbon steel in 10-5 pH lithium hydroxide solution and of Cu-Ni alloys and sacrificial Al-Zn-Sn alloys in synthetic sea-water.

Keywords. X-ray photoelectron spectroscopy; surface film; passivation; inhibition; pitting; sacrificial anodes.

PACS No. 81.60

1. Introduction

Most of the engineering metals and alloys possess a high degree of corrosion resistance due to the presence of surface films which are often adherent, tenacious and resistant to attack by the environment. The surface properties are very important as every type of reaction, be it solid-solid, solid-liquid or solid gas, commences at the surface. For example austenitic stainless steels are resistant to attack due to the presence of a chromium rich passive layer. Ni-Cr alloys (15 to 20% Cr) possess high temperature oxidation resistance due to the formation of a protective Cr₂O₃ or NiCr₂O₄ scale. Further, corrosion inhibition studies have indicated that inhibitors often change the surface film to provide better protection. A mention may be made of the formation of copper-benzotriazole complex on copper and copper-nickel alloys in the presence of benzotriazole inhibitor in an aqueous medium.

The study of the surface composition and property correlations provides basic guidelines to understand the various types of corrosion failures like localised attack, intergranular corrosion, stress corrosion cracking etc. Though other conventional methods of corrosion evaluation like weight change, electrochemical measurements etc have been in use, x-ray photoelectron spectroscopy (XPS) has emerged as one of the most powerful tool for the study of composition of 2–3 nm thick corrosion films. The technique was first developed by Siegbahn (1967) who had earlier shown the possibility of distinguishing from the photoelectron spectrum, an element in different valence state. Since then, XPS has been utilised in a variety of research problems in a number of laboratories. In XPS, all the energy of the incident photon is expended in the ejection of
the photoelectron. Thus, the binding energy \( E_B \) of the electron with respect to Fermi level can be determined by the relationship:

\[
h v = E_{\text{photo}} + E_B + Q,
\]

where \( E_{\text{photo}} \) is the kinetic energy of the photoelectron and \( Q \) is the work function of the spectrometer material. The energy of the photon, \( hv \), is a known quantity. \( E_B \) assumes different values for different electronic shells and it is to some extent, dependent on the chemical state of the atom. Consequently, the energies of the ejected photo-electrons may be shifted slightly as the chemical state is changed. Shifts in core binding energies can be measured and related to the chemical environment. Studies of the satellite structure in the spectra, measurement of relative intensities as well as energies of photoelectron peaks and determination of the angular distribution of the expelled photo-electrons could provide useful information about the chemical species. XPS has thus attracted the attention of the corrosion scientists, as the technique is capable of analysing few monolayer thick films, which cannot be analysed by other techniques. Thus, the action of inhibitors and the phenomenon of passivation can now be understood in a much better way.

As already mentioned, XPS spectrum of an oxide can be distinguished from that of the parent metal. In many cases, it is possible by observing the chemical shifts (Castle and Epler 1974; Hoare and Talerman 1972; McIntyre and Zetaruk 1977; Asami and Hashimoto 1977; Allen et al 1974), peak shapes (McIntyre and Zetaruk 1977), multiplet splitting (Carver 1972) and presence or absence of satellites (Castle 1971). Hydroxides (McIntyre et al 1976) and chlorides (Belton and Clarke 1976; Kishi and Ikeda 1974) can also be identified. Carbon in carbide form can also be distinguished from that due to hydrocarbon contamination (Ramqvist 1970). One difficulty that frequently arises in XPS characterization of corrosion products is that the oxide film formed in aqueous solution on certain metals and alloys may be susceptible to further atmospheric oxidation (before introducing the specimen into the vacuum chamber of the spectrometer) and one may end up with erroneous conclusions. In addition, valence state determination of the oxide is meaningless after \( \text{Ar}^+ \) ion bombardment unless the oxide has been checked for stability.

2. Brief review of corrosion research using XPS

2.1 Passivation studies on iron and its alloys

Several studies have been reported on the passivation of iron and iron-chromium alloys. Roberts and Wood (1977) have investigated the mechanism of oxidation and passivation of iron by water vapour using XPS. The authors have studied the interaction of water vapour with iron at 290 K by XPS and have found clear evidence for the passivation of surface. The O 1s peak observed at 530 eV has been assigned to the oxygen chemisorption or a surface oxide while that at 532 eV to the OH (adsorbed) species. The peak at 534 eV has been related with the molecularly adsorbed water. After having established the presence of both OH\(^{\cdot}\) and O\(^{\cdot}\) species, the authors have suggested that the first step is the dissociation of water followed by interaction between the surface hydroxyls to give chemisorbed oxygen:

\[
\begin{align*}
\text{H}_2\text{O} & \rightarrow \text{OH} (\text{ads}) + \text{H} (\text{ads}) ; \\
2\text{OH} (\text{ads}) & \rightarrow 2 \text{O}(\text{ads}) + \text{H}_2 (\text{g}).
\end{align*}
\]
A significant feature of the study was that the ratio of the intensities of the 530 and 532 eV peaks remained similar up to the maximum water exposures. It was therefore suggested that a discrete chemical phase of FeOOH existed on the surface as a result of passivation.

XPS studies on the electrochemical passivation of chromium (Bouyssoux and Romand 1977) in 0·5 M H₂SO₄ at +0·55 V/NHE for 90 minutes indicated two spin doublets Cr 2p₁/₂ and Cr 2p₃/₂. The higher binding energy of 2p₃/₂ peak (575·6 eV) has been assigned to the oxide while lower binding energy (573·1 eV) to the metal. These peaks have been compared with those of standard oxides of chromium and it has been concluded on the basis of O 1s peak at 529·4 eV that the formation of anhydrous phase of Cr₂O₃ takes place as a result of passivation.

Asami et al (1978) have reported the composition of passive film formed on an extremely corrosion-resistant amorphous alloy of iron (Fe-10 at. % Cr-13 at. % P-7 at. % C), in 1 N HCl as a result of passivation by immersing for 168 hours at 30°C. The study indicated that the air-formed film on the alloy mainly consisted of ferric oxy-hydroxide. However, on immersing in HCl, chromium peak at 577 eV dominated and the traces of iron in the film existed both in Fe²⁺ and Fe³⁺ states. The chromium peak at 577 eV has been attributed to oxidised chromium. The low binding energy peak of O 1s spectrum of the air-formed film, which has been assigned to large amount of OH groups, decreased in intensity by immersing into HCl. The authors have assigned this observation to the formation of a homogeneous, hydrated chromium hydroxide CrOₓ(OH)ₙH₂O and thus causing passivation of the surface. Asami et al (1978) have further investigated the passivity of a series of iron-chromium alloys in sulphuric acid reported that chromium content of the passive film increased drastically at ca 13 at. % Cr of the bulk alloy composition whereas no composition change occurred in the surface of the substrate alloy immediately under the passive film. The high corrosion resistance of iron-chromium alloys with high chromium content has been attributed to the protective nature of hydrated chromium oxy-hydroxide which is the main constituent of passive films in alloys containing chromium at least 12·5 at. %.

Chance and Gaarenstroom (1980) have characterised the surface films formed on a ferritic stainless steel (SAE type J 409) in sodium sulphate solutions of differing pH as a function of polarization potential. The study showed that the critical current for passivation was extremely sensitive to pH, e.g. with 0·5 M sulphate solutions at 82°C, the critical current for passivation increased by more than two orders of magnitude as the pH changed from 3·0 to 2·8. XPS investigations indicated that composition and thickness of a surface film changes significantly during anodic polarization but the most noteworthy was the change from the films of iron and oxygen in the prepassive region into the chromium and oxygen films in the passive region. The films which were initially made of Cr and O did not change above primary passivating potential. The authors have reported thickening of films in the prepassive region during anodic polarization and then thinning of the films on attaining passivity. It has been observed that the film thickness on alloy surfaces at open circuit potential is inversely related to the corrosivity of the solution i.e. film thickness decreases as the corrosivity of the solution decreases. Film thicknesses in the above investigation have been determined by AES depth profiling technique, the discussion of which is outside the scope of this paper.
2.1a Studies conducted in BARC  Correlation between the corrosion behaviour and composition of the oxide film formed on carbon steel in 10-5 pH lithium hydroxide solution has been established. Carbon steel, when exposed to deaerated 10-5 pH lithium hydroxide solution at room temperature showed negligible corrosion rate. The film formed after 24 hours was examined by XPS. The studies in this section and elsewhere reported in the paper were carried out using PHI model 551 electron spectrometer with MgKα source. The instrument was calibrated with respect to Au 4f7/2 peak at 83.8 eV. C 1s peak at 285.0 eV due to hydrocarbon contamination was used to account for the charging effects. High resolution spectra were recorded using PDP-11/04 computer interfaced with the system. Chemical states of the metal cations in the surface oxide films were determined by comparing positions, shapes and shifts of the peaks of different elements with those of the standard spectra. In Fe 2p spectrum (figure 1), Fe 2p3/2 and Fe 2p1/2 peak maxima occurred at 710.8 eV and 725.1 eV (after correction for surface charging). Fe 2p3/2 peak for Fe2O3 has been reported to shift towards higher binding energy by 4.2 eV (Gimzowski et al 1977; Yin et al 1972) with respect to pure metal (706.8 eV) while a shift of 3.0 eV has been estimated for Fe3O4 (Ertl and Wandelt 1975; Coad and Cunningham 1974). Kishi and Ikeda (1973) during oxidation studies of iron observed a maximum shift of 4.0 eV.

Allen et al (1974) have reported Fe 2p3/2 peaks for α-Fe2O3 and α-FeOOH to appear at 711.4 and 711.0 eV respectively while Asami et al (1977) report the Fe 2p3/2 peak for α-Fe2O3 at 710.97 eV, for α-FeOOH at 711.44 eV and for γ-FeOOH at 711.60 eV. McIntyre and Zetaruk (1977) report the values of 711.0 ± 0.15 eV for α- and γ-Fe2O3 and 711.9 ± 0.2 eV for α-FeOOH. On the basis of these reports, it can be concluded that

![Fe 2p spectrum for carbon steel exposed to deaerated 10-5 pH lithium hydroxide solution.](image-url)
a shift of 4.0 eV observed by us of the Fe 3\(p_{3/2}\) peak with respect to pure metal indicates iron to be in Fe (III) state.

Figure 2 shows the O 1s spectrum which appears to consist of two peaks of comparable intensities. The peak at 530.0 eV arises as a result of metal-oxygen bond (hereafter will be referred to as OM peak) and that at 531.6 eV arises as a result of metal-hydroxide bonds (referred to as OH peak later in the text). Asami et al (1977) have reported OM and OH peaks for \(\alpha\)-and \(\gamma\)-FeOOH at 530.1 eV and at 531.4 eV respectively while McIntyre et al (1977) report OM peak for \(\alpha\)-FeOOH at 530.3 \(\pm\) 0.2 eV and OH peak at 531.4 \(\pm\) 0.2 eV. On the basis of these reports, it can be concluded that the oxide film formed in this investigation consists of FeOOH, which protects carbon steel from corrosion in deaerated 10-5 pH lithium hydroxide solution. However, it is difficult to say whether FeOOH formed on the surface is in \(\alpha\) or \(\gamma\) form.

2.2 Inhibition studies

Inhibition studies using XPS have mainly been reported on copper, iron and nickel. Lewis and Fox (1978) estimated the thickness of inhibitive films formed on copper surfaces immersed in solutions of benzotriazole (BTA) by three methods based on XPS and compared with the results obtained using a fourth method viz electrical capacitance of the films. The methods based on XPS involve (a) measurements of the rate of removal of the film by argon ion sputtering, (b) observation of the attenuation of the substrate Cu 2p photoelectron signal on passing through the protective film and (c) a calculation of film thickness from measurements of the intensity of N 1s photoelectron signal from the surface film.

![Figure 2. O 1s spectrum for carbon steel exposed to deaerated 10-5 pH lithium hydroxide solution.](image)
In procedure (a), the authors have used the expression

$$J. S. \frac{t}{\rho} = I. S. \frac{t}{A. e. \rho},$$

(2)

for the removal rate of surface molecules by argon ion sputtering where $J$ is the ion flux, $I$ is the current, $S$ is the sputtering yield (i.e. the number of surface molecules removed per incident argon ion), $A$ is the total area bombarded, $e$ is the electronic charge, $\rho$ is the density of molecules in surface and $t$ is the thickness of a monolayer. The accuracy of this method depends upon the value of $S$. In the absence of measured values for this quantity, the direct comparison of heats of sublimation of two materials for at least one of which $S$ is known, would give a satisfactory estimate for the other. Method (b) employed earlier by Carlson and McGuire (1972) is useful when the film and substrate contain a common element, as is the case for all oxide films formed in corrosion processes. Photoelectron intensity from the substrate is assumed to decrease exponentially with thickness of overlying film, then

$$I_s = I_s^0 \exp \left(-\frac{d}{\lambda \sin \theta}\right),$$

(3)

for electrons emerging at an angle $\theta$ from the surface. $I_s^0$ is the intensity without the film, $d$ is the film thickness and $\lambda$ is the attenuation length (or mean escape depth) for photoelectrons of a particular energy. In this method, the greatest uncertainties lie in assuming that the dynamics of photoemission and emergence of the electrons produced are similar in all materials used in making the measurements. In method (c) the authors have made use of the fact that nitrogen was such an element which occurred only in the film. The absolute intensity of photoelectrons emitted from N atoms has therefore been correlated with the film thickness. The authors claim to have achieved close agreement between the values obtained by these three methods for determining the film thickness.

Surface films formed by two structurally related corrosion inhibitors viz 2-mercapto benzothiazole (MBT) and 2-mercapto benzimidazole (MBIA), on copper studied by XPS and X-ray induced Auger spectroscopy have been reported by Chadwick and Hashemi (1979). In this work the films were formed in 0–0.5 NaCl solutions of pH 2–7 on two types of copper samples—one mechanically polished with a thin Cu₂O film and the other with thick Cu₂O film grown electrochemically in KOH. The presence of MBT was established by S 2p and N 1s peaks in the XPS spectra. The measured Cu 2p binding energies and the lack of Cu 2p shake-up satellites in the spectra represented copper to be either in Cu(O) or Cu(I) state. The presence of MBT on copper surface did not prove that there was compound formation involving copper, and this evidence was obtained by features of CuL₃M₄½M₄½ Auger peak. According to the mechanism proposed for film formation, Cu₂O first dissolves in acid conditions followed by precipitation of Cu-inhibitor complex on to the metal surface, the thickness of the inhibitor film being low on the copper sample with thin Cu₂O film. It has therefore been suggested that Cu ions originating from the Cu₂O dissolution are involved in the precipitation process. The authors have supported this suggestion by further experimental evidence. Other important studies on inhibitor films which may be of interest to the readers have been reported by Wood and Vannerberg (1978) and Kishi et al (1979).

2.2a Studies conducted at BARC on Cu-10 Ni alloy The corrosion rate of pure copper in synthetic sea water was 12.4 mdd while that of Cu-10Ni alloy was found to be 3.2 mdd. Ni alloy when exposed to synthetic sea water polluted with 40 ppm sodium sulphide corroded at a rate of 5.1 mdd. All these rates were evaluated on the basis of 15 days
exposure. XPS investigation on these samples were carried out to find out the reason for the decrease in corrosion rate of copper on alloying with nickel and for the increase in the corrosion rate of this alloy in polluted sea water as compared to that in unpolluted sea water.

The film formed on pure copper in synthetic sea water was found to be Cu$_2$O in accordance with the observation of other workers (North and Pryor 1970; Popplewell et al 1973; Kato et al 1980). XPS survey scan for Cu-10Ni alloy exposed to unpolluted synthetic sea water (figure 3) indicated that main elements in the film were Cu, O, C and traces of Cl. High resolution Cu 2p spectrum (figure 4) showed no satellites, with 2p$_{3/2}$ peak appearing at 932.7 eV and 2p$_{1/2}$ peak at 952.6 eV. LMM Auger peak maximum for copper (figure 5) occurred at a kinetic energy of 916.6 eV. O 1s peak was asymmetric with a maximum at 530.7 eV (figure 6) and with two higher binding energy shoulders e.g. at ca 533 eV and ca 535 eV. No appreciable signal from nickel was obtained from the outermost film.

It is now established that the shake-up satellite structure is closely related to the oxidation state of copper (Larson 1974; Evans 1975; Rosencwaig and Wertheim 1973; Frost et al 1972; McIntyre and Cook 1975). All cupric compounds whose Cu 2p spectra have been studied show satellite peaks believed to arise due to multi-electron excitation process whereas Cu 2p spectra of cuprous compounds do not show satellites. Satellites observed by Novakov (1971), Novakov and Prins (1971) and Schoen (1973) in cuprous compounds could be due to surface contamination by cupric species as the cuprous compounds are susceptible to atmospheric oxidation into cupric (Roberts 1974).

Wallbank et al (1973) suggested with evidence that satellites originate in 2p spectrum from simultaneous transition to 3d states and if 3d states are fully occupied (as in cuprous species), transition to them are not possible. On the basis of the above

![Figure 3](image-url)  
**Figure 3.** XPS survey for Cu–10Ni alloy exposed to synthetic sea water for 15 days.
Figure 4. Cu 2p spectrum for Cu–10 Ni alloy exposed to synthetic sea water; and synthetic sea water + 40 ppm Na₂S for 15 days.

Figure 5. Cu LMM Auger spectrum for pure copper and Cu–10 Ni alloy exposed to synthetic sea water; synthetic sea water + 40 ppm, Na₂S for 15 days.
Figure 6. O 1s spectrum for Cu–10 Ni alloy exposed to synthetic sea water; synthetic sea water after sputtering; and synthetic sea water + 40 ppm Na$_2$S for 15 days.

discussion it can be concluded that copper in this film is not in cupric state. LMM Auger peak at a kinetic energy of 916.6 eV which is quite different from that for pure copper (at 918.8 eV) confirms the presence of cuprous state (McIntyre et al. 1976). O 1s peak at 530.7 eV can be attributed to the oxide oxygen in Cu$_2$O (McIntyre et al. 1976; Evans 1975). Higher binding energy shoulders are due to hydroxide and molecular water which disappear on sputtering. The difference between the binding energies of Cu 2p$_{3/2}$ and O 1s peaks is 402.0 eV which agrees well with the values reported for Cu$_2$O (Evans 1975). Hence, outermost corrosion film in this case can be concluded to consist of Cu$_2$O with traces of chlorine in the film.

There was no significant change in copper spectrum on sputtering, but additional nickel signal was obtained. Ni 2p$_{3/2}$ peak showed two maxima e.g. at 854.5 eV and at 856.4 eV. There was a shoulder at ca 852.8 eV and shake-up satellite structure at ca 861 eV and at ca 864 eV (figure 7). XPS investigations of nickel-oxygen system reveal a shift of 1.6 eV for Ni 2p$_{3/2}$ peak in NiO with respect to that for pure metal (McIntyre et al. 1976; Kim et al. 1974; Brundle 1975; Allen et al. 1979). The peak at 854.5 eV and a satellite at ca 861 eV agree well with the results of Kim et al. (1974) for NiO but the peak at 856.4 eV is too intense to be confused as multiplet splitting of the peak at 854.5 eV (McIntyre and Cook 1975; Kim and Winograd 1974; Kim et al. 1974). It appears that the peak at 856.4 eV and the satellite at ca. 864 eV occur as a result of higher oxidation state of nickel. O 1s peak after sputtering became symmetric and the maximum occurred at 530.4 eV which is not unexpected because the major portion of the film still consisted of
Cu$_2$O. Haber et al (1976) have reported the binding energy values of O 1s peak associated with Ni$^{3+}$ and Ni$^{2+}$ as 529.3 eV and 529.8 eV respectively; however, other authors (Kim and Winograd 1974; Brundle 1975; Norton et al 1977) report the value associated with Ni$^{3+}$ as ca 531.5 eV. It can therefore be concluded that the inner film also consisted mainly of Cu$_2$O with nickel ions in two different valence states viz Ni$^{3+}$ and Ni$^{2+}$ either as a dopant in Cu$_2$O film or simultaneously existing as a mixture of NiO and Ni$_2$O$_3$, however, the O 1s spectrum does not provide evidence to the presence of separate nickel oxide because of domination of Cu$_2$O in the corrosion film.

The film formed on copper in synthetic sea water was Cu$_2$O which is a p-type semiconductor while that formed on Cu-10Ni alloy in the same medium was Cu$_2$O doped with Ni$^{2+}$ and Ni$^{3+}$. Doping by nickel ions leads to an increase in ionic and electronic resistivity and consequently the corrosion rate of the alloy decreases in comparison to that of pure copper (Popplewell 1978).

The film formed on Cu-10Ni alloy in polluted sea water was uniformly black and survey scan indicated the presence of Cu, O, S and C. Cu 2$p_{3/2}$ peak appeared at 932.4 eV (figure 4) while LMM Auger peak occurred at a kinetic energy of 916.9 eV (figure 5) indicating copper to be present in cuprous state. O 1s showed a peak with two maxima viz 532.0 eV and 532.6 eV and three shoulders at ca 530.7 eV ca 533 eV and ca 535 eV (figure 6). The peak at 535 eV represented molecular oxygen while that at 532.6 eV and shoulder at ca 533 eV represented oxygen in the water molecules bound to the surface film. The shoulder at ca 530.7 eV occurred due to Cu$_2$O film on the surface. S 2p peak appeared at 162.0 eV (figure 8) with a higher binding energy shoulder at ca. 163 eV and shake-up satellite structure at ca 167.5 eV and ca 169.6 eV. This peak can be attributed to sulphide species (Lindberg et al 1970; Binder 1973). Thus, the corrosion product in this case was a mixture of Cu$_2$O and Cu$_2$S with adsorbed water molecules and oxygen. Cu$_2$S is known to be a more defective structure than Cu$_2$O, and Cu$_2$O/Cu$_2$S corrosion product film is much more defective than Cu$_2$O alone, thus resulting in an increased ionic and electronic transport through the film (Bates and

![Ni 2p spectrum for Cu–10 Ni alloy exposed to synthetic sea water for 15 days after sputtering.](image)
2.3 Pitting studies

Sugimoto and Sawada (1977) have studied the effect of molybdenum additions to stainless steels in inhibiting pitting in acid chloride solutions. Solution annealed austenitic 20Cr-25Ni steels containing 0–5 wt% Mo were polarised in deaerated 1N HCl (pH 0.02) in steps of 25 mV from the corrosion potential and xps investigations were carried out at different stages of polarization. Anodic polarization of pure Fe, Cr, Ni and Mo showed that Fe and Ni did not passivate and MO exhibited transpassive anodic dissolution above $-0.03$ V and consisted a film of MoO$_3$ above $-0.03$ V. Anodic polarization of Fe-Cr alloys containing 15–70 wt% Cr indicated no pitting in the alloys containing $>40$ wt% Cr. On alloying with 5 wt% Mo the amount of Cr needed to inhibit pitting was reduced to half of those without Mo. xps investigations on the passive film formed on 20Cr-25Ni-5Mo steels showed the film to be an oxyhydroxide which contained Fe$^{3+}$, Cr$^{3+}$, Ni$^{2+}$ and Mo$^{6+}$ as cations. The authors have concluded that Cr$^{3+}$ and Mo$^{6+}$ play a very important role in improving the corrosion resistance of the films. The formation of very compact surface films of CrOOH containing cations like Mo$^{6+}$, Fe$^{3+}$ and Ni$^{2+}$ has been concluded to be the reason for increase in pitting resistance of Cr-Ni steels on addition of Mo.

Figure 8. S 2p spectrum for Cu–10 Ni alloy exposed to synthetic sea water + 40 ppm Na$_2$S for 15 days.
2.4 Studies on sacrificial anodes for cathodic protection

Studies on the xps investigations on aluminium sacrificial anodes have been carried out in authors' laboratory and have been reported in detail elsewhere (Pai et al 1982). Pure aluminium is not recommended for sacrificial anodes because of passive oxide film on its surface. Judicious additions of Zn, Mg, Cd, Hg, Sn and In as alloying elements can possibly modify this oxide film on aluminium and render Al anodes suitable for cathodic protection of steel structures in sea water.

In this investigation, studies were carried out on Al-5Zn-X Sn alloys where X varied from 0.67% to 0.5%. Current efficiencies (%) of alloys with different heat treatments were measured after coupling to mild steel in sea water for 180 hr. Alloy Al-5Zn-0.1 Sn (heated for 3 hours at 300°C and water quenched) was found to behave best as a sacrificial anode while the same alloy after heating for 3 hr at 300°C and furnace cooling, was not found to be so good. On exposing these specimens to sea water, xps studies indicated that the film formed on the quenched alloy contained, in addition to Al$_2$O$_3$, relatively greater amount of ZnO than that formed on the furnace cooled alloy. In the oxide layer where ZnO and Al$_2$O$_3$.3H$_2$O are formed, Al$^{4+}$ (ionic radius 0.51 Å) ions act as dopant in ZnO film (ionic radius of Zn$^{2+}$ 0.93 Å) and increase the cation vacancy concentration. The higher oxygen potential gradient developed due to reduced aluminium activity favours the higher rate of dissolution of Al alloy in sea water by providing driving force for the movement of cations from alloy-oxide interface. Obviously, the quenched alloy in which the amount of ZnO formed is more, will behave as a better sacrificial anode.

3. Conclusions

From the above illustrations, it can be seen that xps is a very powerful technique for understanding and solving the corrosion problems. It has been clearly demonstrated that xps can be used for determining the nature, thickness and composition of passive films, for inhibition and pitting studies and for understanding the phenomenon occurring in the cathodic protection. The authors' studies indicate that formation of FeOOH provides protection to carbon steel against corrosion in deaerated alkaline medium, better corrosion resistance of Cu-Ni alloys in sea water is due to the formation of Cu$_2$O film doped with Ni$^{2+}$ and Ni$^{3+}$, the enhancement of corrosion of these alloys in sulphide polluted sea water occurs due to the formation of porous corrosion product Cu$_2$O/Cu$_2$S and sacrificial aluminium-zinc-tin alloy anodes behave better when the oxide film is rich in zinc.

References

Asami K and Hashimoto K 1977 Corros. Sci. 17 559
Asami K, Hashimoto K, Masumoto T and Shimodaira S 1976 Corros. Sci. 16 909
Asami K, Hashimoto K and Shimodaira S 1978 Corros. Sci. 18 151
Bates J F and Popplewell J M 1975 Corrosion 31 269
Belton P S and Clarke T A 1976 Philos. Mag. 34 157
Bouyssoux G and Romand M 1977 J. Electron Spectrosc. Relat. Phenom. 11 185
Brundle C R 1975 Surface Sci. 48 99
Carver J C 1972 in Electron spectroscopy (ed.) D A Shirley (Amsterdam: North Holland) p. 781
Chadwick D and Hashemi T 1979 Surface Sci. 89 649
Chance R L and Gaarenstroom S W 1980 Corrosion (NACE) 36 94
Coad J P and Cunningham J G 1974 J. Electron Spectrosc. Relat. Phenom. 3 435
Ertl G and Wandelt E 1975 Surface Sci. 50 479
Evans S 1975 J. Chem. Soc. (Faraday Trans. II) 71 1044
Kim K S and Winograd N 1974 Surface Sci. 43 625
Kishi K, Chinomi K, Inoue Y and Ikeda S 1979 J. Catal. 60 228
Lewis G and Fox P G 1978 Corros. Sci. 18 645
McIntyre N S and Cook M G 1975 Anal. Chem. 47 2208
McIntyre N S and Zetaruk D G 1977 Anal. Chem. 49 1521
Novakov T 1971 Phys. Rev. B3 2693
Novakov T and Prins R 1971 Solid State Commun. 9 1975
Roberts M W and Wood P R 1977 J. Electron Spectrosc. Relat. Phenom. 11 431
Roberts R F 1974 J. Electron Spectrosc. Relat. Phenom. 4 273
Rosencwaig A and Wertheim G K 1973 J. Electron Spectrosc. 1 493
Schoen G 1973 Surface Sci. 35 96
Wood J C and Vannerberg N-G 1978 Corros. Sci. 18 315
Yin L I, Ghose S and Adler I 1972 Appl. Spectrosc. 26 355
Irradiation behaviour of nuclear fuels

P R ROY and D N SAH
Radiometallurgy Division, Bhabha Atomic Research Centre, Trombay, Bombay 400085, India

Abstract. This paper gives a brief review of the important phenomena observed in metallic uranium and ceramic nuclear fuels during irradiation in reactors. The mechanism of irradiation growth, irradiation creep and swelling which are responsible for the dimensional instability of uranium has been described. Important phenomena observed in ceramic nuclear fuels, e.g. fuel densification, fuel restructuring, plutonium segregation, oxygen and fission product migration, irradiation creep, fission gas release and swelling have been discussed. A brief note is included on computer modelling for prediction of fuel element irradiation behaviour.

Keywords. Irradiation effects; nuclear fuels; metallic fuels; ceramic fuels; irradiation behaviour; irradiation growth; irradiation creep; swelling; densification; restructuring; fission gas release; fuel modelling.

PACS No. 28-40; 28-90

1. Introduction

Nuclear fuels are exposed to a unique and very severe thermal and radiation environment during their use in reactors. The behaviour of the fuel material in this environment is quite different from its behaviour outside the reactor. It is governed by the complex interplay of a large number of interactive physical, chemical, mechanical and metallurgical processes which become operative in the fuel in the reactor environment. A great deal of studies have been made and are still continuing to understand the irradiation behaviour of fuels in nuclear reactors. These studies have revealed spectacular changes in the geometry, dimensions, composition and microstructure of the fuels during and after fission due to irradiation. In this paper some of the important phenomena observed in two types of nuclear fuels viz metallic uranium fuels and ceramic fuels (oxide, mixed oxide and mixed carbide), which are relevant to India's nuclear programme, are described. As oxides are now the most widely used fuels in the present-day nuclear power reactors, they have been discussed at length. The paper briefly reviews the important findings about these phenomena and their mechanisms.

2. Irradiation damage in solid nuclear fuels

The radiation environment in the nuclear reactor core consists of fission fragments, high energy neutrons, charged particles and other radiations. The main source of radiation damage in nuclear fuels is the highly ionised fission fragments carrying a total of 165 MeV energy and the fast neutrons which have an energy upto 10 MeV. Fission
fragments and neutrons move rapidly through the lattice, exchanging their energy with the lattice atoms until they come to rest. The damage to the fuel by fission fragments is caused in the following three ways:

(i) **Fission spikes**: Each fission fragment is highly ionised (about 20+) and has a range of 6–10 microns in the fuel matrix. Over this small distance, the fission fragment releases all its energy which has an effect of raising the fuel temperature to more than 3000°C for a brief period of $10^{-11}$ seconds. This zone is called a fission spike. It is a cylindrical region of about 100 Å in radius with the fission fragment track as its axis. The end portion of the fission spike where the fission fragment comes to rest is characterised by severe disarrangement of lattice atoms and is often called displacement spike while the rest is called thermal spike (figure 1). Once the fission fragments come to rest, they are more usually described as fission products.

(ii) **Lattice defects**: Collision of fission fragments with lattice atoms of the fuel matrix produces primary knock-on atoms which have sufficient energy to displace other atoms from the fuel lattice. This gives rise to creation of several types of point defects in the lattice (figure 2). The formation of lattice defects in the fuel matrix provides the impetus for atomic diffusion of various species. All diffusion-controlled processes are enhanced.

---

**Figure 1.** Schematic representation of the fission spike in nuclear fuel. The fission fragment enters the lattice at left as shown by the arrow mark (Kopelman 1959).

**Figure 2.** Schematic picture illustrating the formation of different types of lattice defects by a primary knock on atom. The knock-on atom enters the lattice at left and comes to rest at P (Chadderton and Torrens 1969).
Irradiation behaviour of nuclear fuels

and new transport processes are induced at quite low temperatures. The two basic types of lattice defects, namely, the vacancy and the interstitial recombine if the temperature is high. But they remain as single defects at lower temperature and have a large effect on the fuel element behaviour.

(iii) Impurity atoms: Each fission event creates two atoms of fission products which are chemically different from the original material. Presence and build-up of these impurity atoms influence properties of the fuels and a detailed study of the fission products is therefore very important.

3. Irradiation effects in metallic fuel

Uranium metal exists in three allotropic modifications with different crystallographic structure: Orthorhombic alpha uranium up to 668°C, tetragonal beta uranium (668–775°C) and body-centred cubic gamma uranium (775–1130°C). Most metallic uranium fuels are used in the alpha state. Orthorhombic alpha uranium shows highly anisotropic thermal expansion and exhibits poor dimensional stability during irradiation. The dimensional changes in alpha uranium result from three types of irradiation-induced phenomena: irradiation growth, irradiation creep and fission product swelling.

3.1 Irradiation growth

Irradiation growth refers to the change in uranium shape at constant volume without any external stress application. The phenomena of growth were reported by Konobeevsky et al (1955), Paine et al (1955) and Pugh (1955).

During irradiation a single crystal of alpha uranium exhibits growth in ⟨010⟩ direction, shrinkage in ⟨100⟩ direction and remains unchanged in ⟨001⟩ direction. In a polycrystalline sample the extent of dimensional changes depend upon the degree of preferred orientation in the specimen. The growth of uranium is expressed by the following correlation (Robertson 1969):

\[ L = L_0 \exp \left( \frac{G}{f} \right) \]

where \( L \) is the length of specimen after a fraction \( f \) of the atoms have fissioned, \( L_0 \) is the initial length of the specimen and \( G \) is the growth constant. The value of \( G \) denotes the susceptibility of uranium to growth and depends upon uranium texture and irradiation temperature. Growth constant \( G \) has very large values at low temperature (sub-zero temperature) but relatively independent of temperature in the range 0–300°C and small values at high temperatures (Buckley 1964). No growth is observed at temperatures above 500°C.

The phenomena of irradiation growth is connected with the anisotropic thermal expansion of the alpha-uranium (figure 3). The direction of growth in uranium coincides with the direction which has a negative coefficient of thermal expansion. Similarly the direction which shrinks during irradiation corresponds to the direction of highest thermal expansion coefficient. The mechanism of irradiation growth involves formation of extra monolayer plates of lattice defects created by the passage of fission fragments in solid uranium on certain crystal planes (figure 4). The thermal stresses due to anisotropic expansion of uranium in the fission spike, force the vacancies and interstitials to aggregate on different crystal planes to cause elongation (due to interstitial clusters) in the direction ⟨010⟩ and contraction (due to vacancy cluster) in
<100> direction. At higher temperature (above 500°C) the point defects created in uranium are sufficiently mobile to get annihilated and hence no growth is observed. Since growth is enhanced by preferred orientation the only way to control growth is to have a random texture. This is obtained by beta-quenching of uranium which assures fine randomly oriented grains in the fuel.

3.2 Irradiation creep

Irradiation creep and irradiation growth play an important role in the often observed bowing of the uranium fuel element during irradiation (figure 5).
Figure 5. Bowing observed in an experimental uranium fuel element of Dhruva reactor (irradiated in CIRUS).
Accelerated creep of uranium during irradiation was reported by Konobeevsky et al (1955). It was found by Roberts and Cottrell (1956) that in a thermal neutron flux of $10^{12} \text{n/cm}^2/\text{sec}$ at temperature of 100°C, a stress as low as 1% of yield strength caused appreciable creep in polycrystalline uranium. This observation was explained on the basis of yielding creep mechanism (figure 6). In this mechanism creep is caused by build-up of high internal stresses between the grains due to growth of misaligned grain in a polycrystalline material. When the internal stress reaches the yield stress, the material deforms plastically under the influence of an externally applied stress which is small compared to the yield stress. Most of the work is done by internal stress, the external stress merely guides the deformation. The creep rate due to this mechanism is given by the following correlation (Gilbert 1971)

$$\dot{e}_c = \left(\frac{\sigma}{\sigma_y}\right)\dot{e}_g,$$

where $\dot{e}_c$ is irradiation creep rate, $\sigma_y$ is the yield stress, $\dot{e}_g$ is the growth strain rate and $\sigma$ is the applied stress.

The dependence of creep rate on neutron flux and temperature is contained in the growth strain rate.

Another mechanism of irradiation creep in uranium deals with relaxation of elastic stress in the crystal lattice volume temporarily disorganised by a fission spike (figure 7). The creep rate due to this mechanism is given by the following correlation (Dienst 1977).

$$\dot{e}_c = \frac{\sigma}{E} \dot{F},$$

where $\sigma$ is the stress, $E$ is Young's modulus and $\dot{F}$ is the fission rate.

**Figure 6.** Yielding creep mechanism. Irradiation growth of misaligned adjacent grains produces internal stress (Gilbert 1971).
Irradiation behaviour of nuclear fuels

3.3 Swelling

Swelling due to replacement of uranium atom by fission products was calculated by Howe and Weber (1957) who showed that the percentage volume increase in uranium should be about three times the atom percent burn-up. This estimate compares well with the swelling at lower temperatures where gaseous swelling is absent. However, large amount of swelling is observed in high purity uranium in the temperature range 350–600°C. A breakaway swelling (swelling maxima) is observed at about 450°C and the magnitude of swelling is more in high purity uranium compared to the swelling in uranium containing small amount of impurities (figure 8). Swelling in the temperature range 350–500°C is characterised by formation of large (50–200 microns) cavities and tearing at grain boundaries and at twins (Pugh 1955). In temperature range 500–600°C, swelling is associated with formation of small platelike pores crystallographically aligned in planar arrays within uranium grains (Legget et al 1964). The burn-up at which enhanced swelling appears in uranium depends on the temperature of irradiation. At higher temperature of irradiation, the swelling initiates at low burn-up but when the temperature is low the breakaway swelling can be delayed upto a high burn-up. External restraint is found to suppress the formation of cavities and aligned pores, reducing fuel swelling. Minor addition of alloying elements Fe, Al, C, Si etc in uranium inhibits the swelling.

Metal fuels operate at lower fractions of their melting points where fission gas diffusion is fairly slow. Moreover, the thermal gradient across the fuel is small due to higher thermal conductivity and the driving force for directed bubble migration is small. Bubbles generally remain quite small and gas release is negligible. The swelling of uranium originates from the nucleation and growth of bubbles of fission gases xenon and krypton which are insoluble in fuel matrix and tend to precipitate as bubbles on
grain boundaries, dislocation and twins. However, the anisotropic growth of uranium is considered to be the main driving force (in temperature range 350–500°C) for the formation of cavities and tears which are responsible for large swelling in uranium. The cavities are nucleated at the existing fission gas bubbles. Alternatively the cavities can be nucleated mechanically by the intergranular stresses arising from anisotropic growth of adjacent grains. When the gas bubbles or grain boundary tears exceed a critical size (2000 Å), they rapidly grow to form large cavities. The growth occurs by migration of vacancies to the cavity under the influence of tensile stress caused by intergranular interaction resulting from anisotropic growth of uranium. In the temperature range 500–600°C where irradiation growth is negligible, the swelling is believed to be caused by homogeneous nucleation of voids by condensation of vacancies (McDonell 1973). The suppression of swelling by addition of alloying elements is attributed to interaction of point defects with alloy constituents.

At one time it was thought that the swelling of metal fuels was sensitive to the fuel strength and therefore could be reduced by making stronger fuels with addition of alloy elements like molybdenum. It was, however, recognised later that the main restraining force on gas bubbles was surface tension rather than the matrix strength. Hence, keeping the gas bubbles smaller e.g. by providing numerous nuclei on which they can form rather than strengthening the matrix, could be more effective.

4. Irradiation effects in ceramic fuels

In addition to the intense radiation field, the ceramic fuels are exposed to very severe thermal conditions. Volumetric fission heat generation in ceramic fuels gives rise to large temperature difference between the centre line and external surface of the fuel elements. A typical mixed oxide fuel operates with a high centre temperature approaching melting point and there exists a radial thermal gradient as high as
Irradiation behaviour of nuclear fuels

Figure 9. Changes observed in ceramic fuels during irradiation due to severe thermal and radiation environment in the fuel pellets (Roy and Sah 1983).

3000–5000°C/cm. Besides causing cracking and distortion of the fuel pellets as a result of thermal stress, these unique irradiation conditions provide the driving force for several transport processes in fuel which lead to significant changes in the microstructure, composition, dimension and shape of the fuel. The main irradiation effects (figure 9) can be divided in the following broad categories: (a) densification; (b) restructuring; (c) redistribution of fuel and fission products; (d) effects of fission gases: swelling and gas release; (e) irradiation creep

4.1 Irradiation-induced densification

The problem of irradiation-induced densification which results in formation of large axial gaps in the fuel column and leads to collapse of cladding tube in these gaps, was first observed in 1972 in large pressurised water reactors (pwrs). This sent a shudder through the reactor engineers who had to derate the reactor power pending investigations into the mechanism and possible solution of fuel densification under irradiation. Fortunately, within a short period the mechanism was correctly identified which defined the requisite microstructure (consisting of large grains and controlled porosity) of a densification resistant fuel. Use of densification resistant fuel along with internal prepressurisation of the fuel element provided a satisfactory solution to the problem.

The densification of ceramic fuels during irradiation is caused by fission spike-pore interaction which results in shrinkage and dissolution of pores present in the fuel matrix (figure 10). It takes place quite early in life and has much greater technological significance. Irradiation-induced pore shrinkage and annihilation was observed by Wapham (1966) using electron microscope. Later Ross (1969), Bellamy and Rich (1969) and Turnbull and Cornell (1970) reported similar observations. A fission fragment traversing through fuel grain disrupts the pores on its track. The pore is transformed into lattice vacancies in this process. The vacancies generated in this event migrate to the nearby pores, or to the grain boundaries. Vacancy migration to the grain boundary results in densification of the fuel. Due to large number of vacancies generated by one
fission event and large fission rate ($10^{13}$ fissions/cc), the rate of densification is considerable.

Assman and Stehle (1978) derived an equation for volume change in fuel due to densification. They considered two populations of pores in the fuel body, coarse pores which are relatively stable and fine pores ($< 0.1 \mu m$) which are atomised in a single encounter with fission spike. Assuming densification to occur by flux of vacancies from grain body to the grain boundaries they gave a correlation of the following form for the volume change ($\Delta V/V_0$) in the fuel:

$$\frac{\Delta V}{V_0} = -f^* P_0 \left[ 1 - \exp \left( -n^* \Omega \frac{F_t}{r_g} \right) \right],$$

where

$$f^* = \frac{1}{1 + (4\pi/15) Z_c r_c r_g^2},$$

and

$$n^* = \left( 1 - \left( 4\pi/45 \right) Z_p r_p r_g^2 \right) n.$$

In the above correlation $\Omega$, is volume of fission spike, $F_t$ is burn-up in fission/cc, $n^*$ is the fraction of vacancies per encounter that can escape from the fine pores. The factor $f^*$ ($\leq 1$) describes the effect of coarse pores which capture a fraction of vacancies produced. $r_c$ is radius of coarse pores, $r_g$ is grain size (radius), $Z_c$ and $Z_p$ are densities of coarse and fine pores respectively, $P_0$ is initial porosity. For very small initial concentration of fine pores in the fuel body, $f^* = 1$ and $n^* = n$. The above correlation shows an exponential law for pore removal from fuel and it also brings out the important parameters which have significant effect on in-pile fuel densification.

Figure 10. Mechanism of in-pile fuel densification (Chubb et al 1975).
4.2 Fuel restructuring

The microstructure of the unirradiated sintered ceramic fuel pellet consists of the equiaxed grains of 5–15 micron diameter with uniformly distributed porosity. The cross-section of irradiated fuel however, reveals concentric zones of altered microstructures (figures 11 and 12) which are indicative of the temperature in the fuel during irradiation. These zones have fairly-defined temperatures associated with their boundaries. The different microstructural zones and the temperature range in which they are formed in mixed oxide and mixed carbide fuels are listed in table 1. Equiaxed grain growth law is given by the following correlation (Olander 1976):

\[ D^m = D_0^m + K_0 t \exp(-Q/RT) \]

where \( D \) and \( D_0 \) are the final and the initial grain size, \( K_0 \) is a constant, \( t \) is time, \( Q \) is activation energy and \( T \) is temperature. The exponent \( m \) and the activation energy were found to be 2.5 and 460 kJ/mol respectively for \( \text{UO}_2 \) in laboratory experiments. Under irradiation conditions, the presence of fission gas bubbles on grain boundary has a retarding effect on grain growth rate. Ainscough et al (1973) suggested the following grain growth law for \( \text{UO}_2 \) during irradiation:

\[ \frac{d(D)}{dt} = k \left( \frac{1}{D_0} - \frac{1 + 2000 \Omega F t}{D_m} \right) \]

Figure 11. The cross-section of the irradiated \( \text{UO}_2 \) fuel element of Tarapur Atomic Power Station (TAPS) revealing formation of a porous central region in the fuel pellet, (magnification 6X). The maximum fuel centre temperature was estimated to be 1250°C (Bahl et al 1979).
Figure 12. Restructuring observed in fuel pellet of Tarapur Atomic Power Station (magnification 5X). The maximum fuel centre temperature was estimated to be 2200°C.

Table 1. Microstructural zones observed in the cross-section of an irradiated ceramic nuclear fuel pellet.

<table>
<thead>
<tr>
<th>Fuel type</th>
<th>Temperature range</th>
<th>Microstructural observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(U, Pu)O₂</td>
<td>below 1100°C</td>
<td>As sintered microstructure</td>
</tr>
<tr>
<td></td>
<td>1100–1300°C</td>
<td>Intergranular porosity</td>
</tr>
<tr>
<td></td>
<td>1300–1700°C</td>
<td>Equiaxed grain growth</td>
</tr>
<tr>
<td></td>
<td>Above 1700°C</td>
<td>Columnar grain growth and central void</td>
</tr>
<tr>
<td>(U, Pu)C</td>
<td>below 1100°C</td>
<td>As fabricated microstructure and very small bubbles</td>
</tr>
<tr>
<td></td>
<td>1100–1400°C</td>
<td>Grain growth, large gas bubbles, bubbles at grain boundaries</td>
</tr>
<tr>
<td></td>
<td>1400–1700°C</td>
<td>Radially elongated grains and pores</td>
</tr>
<tr>
<td></td>
<td>Above 1700°C</td>
<td>Irregular, big rounded pores and central void</td>
</tr>
</tbody>
</table>

where $D_m$ is limiting grain size, $k$ is the grain growth constant, $F$ is fission density and $\Omega$ is atomic volume. Formation of radially oriented columnar grains involves movements of pores up the temperature gradient by an evaporation and condensation mechanism. Fuel on the hotter side of a large lenticular pore evaporates and condenses on the cooler side creating a transport process for the pores to migrate up the thermal gradient. The movement of the pores to the centre of the fuel results in high density columnar grains and the formation of a central hole or 'void'. It can take as little as 24 hours at power for
restructuring to occur. The rate of columnar grain growth is governed by temperature and thermal gradient. The process is important at temperatures where the vapour pressure of the ceramic fuel becomes significant.

4.3 Redistribution of fuel constituents and fission products

4.3a Plutonium redistribution: Plutonium-bearing ceramic fuels initially contain uniformly distributed plutonium. During irradiation, however, plutonium tends to segregate to some preferred locations. In mixed oxide (U, Pu)O₂ fuel the segregation of plutonium is observed to be related to the initial stoichiometry of the fuel. The radial profile of plutonium concentration measured in an irradiated mixed fuel of two different initial stoichiometry, showed that in fuel having an initial O/M of 2.01 the plutonium segregated towards fuel centre while in fuel having an initial O/M of 1.935 the plutonium segregation occurred away from the fuel centre (figure 13). This indicated the possibility of the existence of a critical value of stoichiometry for which no plutonium segregation should occur. However, plutonium will segregate towards periphery or towards centre depending upon the fact whether the initial O/M of fuel was lower or higher than the critical value respectively.

The observed plutonium segregation is successfully explained by vapour transport mechanism which involves preferential evaporation and migration of one of the fuel species. The basis for vapour transport mechanism is migration of lenticular pores towards fuel centre by evaporation and condensation of fuel material. The plutonium segregation is caused by the difference in the U/Pu ratio in the solid and its vapour inside the migrating pore. When the O/M ratio of the fuel is such that congruent evaporation can take place then no plutonium segregation will occur. If the stoichiometry is higher (than the stoichiometry for congruent evaporation) the vapour

Figure 13. Radial plutonium profile observed after irradiation in mixed oxide fuels of two different O/M ratios (Bramman and Powell 1975).
phase in the pore consists almost entirely of UO$_3$ with minor amount of UO$_2$ and PuO$_2$. This leads to preferential migration of uranium from high temperature central region making the centre rich in plutonium. When the O/M of the fuel is lower, the vapour phase inside the pore consists of PuO which leads to depletion of plutonium from the centre. Attempts have also been made to explain the redistribution of plutonium on the basis of solid state thermal diffusion of plutonium (Beiswenger et al 1967; Bober et al 1969). Though it was possible to predict plutonium segregation towards centre, the mechanism of thermal diffusion cannot explain the reversal in the direction of plutonium segregation observed at low O/M ratio.

4.3b Oxygen redistribution: In a mixed oxide fuel containing appreciable amount of plutonium, oxygen also gets redistributed under the influence of radial thermal gradient. The oxygen migration in the fuel is governed by the initial stoichiometry of the fuel. In a hypostoichiometric fuel, the oxygen migrates towards the low temperature side (i.e. towards fuel periphery) while in a hyperstoichiometric fuel the oxygen migration occurs from low temperature to high temperature side (figure 14). In both cases, the fuel surface tends to approach stoichiometry.

The redistribution of oxygen observed in the mixed oxide fuel is explained by two mechanisms (Olander 1976): (i) oxygen transport in gas phase by counter diffusion of CO$_2$ and CO, (ii) thermal diffusion of oxygen vacancies. The mechanism of oxygen transport through gas phase is based on the attainment of a thermodynamic equilibrium among CO$_2$, CO and oxygen in the existing temperature gradient. The gases CO and CO$_2$ are formed from the carbon impurity in the fuel. In a hyperstoichiometric fuel the CO$_2$ diffusing to the hotter zone releases an oxygen atom to the solid and the CO formed diffuses back to the cooler zone where it picks up an

![Figure 14. Redistribution of oxygen in mixed oxide fuel pellet of different initial stoichiometry (Assman and Stehle 1982).](image)
Irradiation behaviour of nuclear fuels

411

An oxygen atom to form CO₂ which repeats the cycle. This happens till a constant CO₂/CO ratio determined by the initial composition of the fuel is achieved.

The above model meets with difficulty when applied to transport in the hypostoichiometric fuel. The very low oxygen potential of hypostoichiometric fuel does not allow complete gasification of available carbon. So CO₂ pressure is very low. Therefore the CO/CO₂ mechanisms would be extremely slow in hypostoichiometric fuels. However, experiments have shown a rapid oxygen redistribution in hypostoichiometric fuels. This indicates that other mechanisms may be responsible. Therefore, a model of oxygen transport by thermal diffusion of oxygen has been proposed. In this model the diffusing species are considered to be single oxygen vacancies. The heat of transport for these species being negative, they migrate towards high temperature regions.

4.3c Fission product redistribution: The fission process in nuclear fuel results in production of atoms of more than 30 elements and about 120 isotopes of these elements have appreciable half lives. Due to steep temperature gradient in the fuel some fission products move away from the location where they are produced. Loss or gain of fission products from the region where they were created, alters many important physical and chemical properties of the fuel locally. The effect of fission products on the fuel behaviour is governed by their physical and chemical state. One of the important consequences of fission product redistribution is the possibility of chemical attack by fuel on the cladding material.

Studies of fission product distribution in irradiated oxide and mixed oxide fuel have provided the following important findings: (i) Fission product elements Zr, Nb, Ce, Pr, Ba and Sr which exists as non-volatile oxide in solid solution with the fuel, do not redistribute. (ii) Noble metals Ru, Rh, Pd and a part of Mo exist in elemental form and migrate significantly. They are observed as white metallic inclusions in the columnar grain zone or as alloy in the central void. (iii) Volatile fission products Cs, Rb, I, Te, Sb and Cd tend to migrate to the cooler regions in the fuel. Gross radial migration of cesium occurs from high temperature central regions to cooler periphery fuel and to the pellet interface locations. Axial migration of cesium to the plenum is also observed. This migration behaviour of cesium is attributed to its high volatility. (iv) In fuels containing appreciable amount of plutonium, the fission yield of noble metals is higher and the yield of oxide forming elements is lower compared to their yield in uranium fuel. Due to this effect, a part of oxygen released by fission, remains in uncombined form in the fuel and increases its oxygen potential.

4.4 Effects of fission gases

A substantial fraction (15%) of the fission products in nuclear fuel consists of rare gases, xenon and krypton, which have extremely low solubility in the fuel matrix. If these gases are released from the fuel body they increase the pressure inside the fuel element subjecting the cladding tubes to stress. On the other hand, if they are retained in the fuel, they precipitate as gas bubbles and cause fuel swelling. The dimensional changes in fuel due to swelling leads to mechanical interaction between the fuel and cladding which often leads to pellet clad interaction failure by stress corrosion cracking (figure 15).

4.4a Fission gas swelling: Swelling of the fuel is attributed to the volume of fuel occupied by the gas bubbles. Bubble formation depends on the mobility of gas (either as
single atom or complexes), the minimum number of gas atoms required to form a stable nucleus and the rate at which lattice vacancies can be supplied to enhance the stability of a nucleated complex. In the irradiation environment existing in the fuel the nucleation and growth of gas bubbles is accompanied by the irradiation induced re-solution of bubbles by fission fragments. It is known that bubbles which are stable out-of-pile are susceptible to re-solution in irradiation environment even at temperature as high as 1200°C.

The nucleation of bubbles occurs heterogeneously on the fission fragment tracks (Turnbull 1971). The nucleation sites are believed to be vacancy-rich segment of the fission fragment tracks. Such nucleation is independent of thermally activated transport properties and is therefore intrinsically unaffected by temperature, which is in agreement with the observations of Cornell (1971) who showed that density of intragranular bubbles is insensitive to temperature. Since two xenon atoms form a thermally stable nucleus, homogeneous nucleation of gas bubbles has also been considered but in this case bubble density will depend on temperature which is opposed to the observations of Cornell.

Bubbles nucleated under irradiation grow by capture of gas atoms and vacancies and then get dissolved by irradiation induced re-solution (Turnbull 1971). Concentration of intragranular bubbles is sustained by fresh nucleation. There is a dynamic equilibrium between formation of gas bubbles by nucleation and growth on one hand and resolution of the bubbles by fission fragments on the other. At temperatures above 1200°C the resolution becomes less effective because of the increased gas atom mobility. The growth of bubbles depends on the temperature (which determines the gas diffusion
coefficient) and burn-up which controls the dissolved gas atom concentration. Because of the re-solution process each bubble has an average life defined by

$$T \approx \left[ \frac{1}{2\pi r^2 F \lambda} \right],$$

where \( r \) is bubble radius, \( F \) is fission rate and \( \lambda \) is range of fission fragment. This indicates that the bubble growth is restricted. However, the observed presence of large gas bubbles in the fuel suggested that bubbles larger in size are more stable because the re-solution is not so effective in destroying large bubbles as it is with small bubbles. Turnbull and Cornell (1970) showed that bubbles with diameter greater than 24 Å have enhanced stability as they cannot be destroyed in a single event of bubble-fission fragment interaction.

The equilibrium number of bubble per unit volume, \( n \) in the fuel is given by

$$n/T = 2Fa,$$

where \( a \) is number of bubbles nucleated per fission fragment, \( T \) and \( F \) have the same meaning as in earlier equation.

The equilibrium quantity of gas \( m \) in intragranular bubbles per unit volume of fuel is given by following correlation (Turnbull 1971):

$$m = \frac{4a}{3\lambda^2b'} \left( 2b'D/\pi F\lambda \right)^{1/4},$$

where \( b' \) is Van der Waals constant for fission gas, \( C \) is dissolved gas atom concentration, \( D \) is gas diffusion coefficient, \( F, \lambda \) and \( a \) have same meaning as earlier. This indicates that the quantity of gas in intragranular bubbles increase as fourth root of the dissolved gas concentration. It is clear that the fraction of gas in the intragranular bubble is invariably a small fraction of the entire gas content. The importance of this fact is that the gas from the matrix migrate to grain boundary as single atom and forms intergranular bubbles. In fact a major part of the swelling observed in fuel is caused by formation and growth of large intergranular bubbles in the process of gas release.

Metallographic observations (Chubb 1972) on \( UO_2 \) irradiated at temperature below 1600°C showed an equiaxed grain structure with extensive porosity along the boundaries where fuel temperatures was sufficiently high. This porosity accounts for the majority of the fuel swelling. Detailed investigations of intergranular fractured surfaces show the characteristics of bubbles lying on grain faces and those situated at grain edges and corners. The latter are observed to be extensively developed and generally occupy a larger volume than the relatively small grain face bubbles.

### 4.4b Fission gas release

Release of fission gases from the fuel to the free volume of the fuel elements is governed by the mobility of the gas as individual atoms or as bubbles in the existing condition during irradiation. The release of fission gases is observed to be strongly dependent on temperature. Other factors which influence the release are grain size, burn-up, fission rate, fraction of open porosity, fuel density and stoichiometry of fuel in \( UO_2 \).

Xenon is found to diffuse in \( UO_2 \) as a complex of constant size. This complex consists of a xenon atom bound with one uranium vacancy and two oxygen vacancies. The diffusion of xenon has been studied by Cornell (1969). It is found that xenon has no significant mobility in \( UO_2 \) at temperatures below 1000°C. Migration of bubbles may occur by several processes. The effective diffusion coefficient for migration of gas bubble at a constant temperature is dependent on its radius (\( r \)) and is proportional to
\( r^{-n} \) (Gittus 1978) where the exponent is dependent on the physical process responsible for bubble movement. The value of \( n \) is 4 for surface diffusion, 3 for volume diffusion and 2 when bubble movement occurs by diffusion of atoms in a vapour phase across bubble filled with equilibrium pressure of insoluble gas. It is found that under the effect of temperature gradient surface diffusion is the dominant mechanism of migration for bubbles with radii less than 16 microns. Bubbles of larger (> 16 micron) size move by vapour transport mechanism. However, no bubble migration is observed at temperatures below 1500°C and the main mode of gas mobility upto 1500°C remains atomic diffusion. The method of predicting gas release in terms of the diffusion coefficient alone is, however, not reliable because of the existence of different types of traps like dislocations, grain boundaries, bubbles, lattice defects, defect clusters, precipitates of fission products. Moreover, sudden changes in gas release with changes in reactor power have been observed which could not be accounted for by the diffusion processes. 'Gas burst' is observed to occur at start-up and shutdown and during power transients.

Two basic mechanisms of gas release operate below 1000°C. They are based on two collision processes, (a) 'recoil' in which a fission event near the surface produces a gas atom which has sufficient energy to escape from the solid, (b) 'knock-out' where a fission fragment collides with a gas atom with sufficient energy to knock it out of the solid.

In the temperature range 1000–1600°C the release occurs by atomic diffusion of gas atoms from grain interior to the grain boundaries (figure 16) where the ready supply of vacancies permits the gas atoms to form gas bubbles large enough to survive the re-solution effect. Bubbles on the grain boundary grow by collection of atoms from grain interior. If the temperature is sufficiently high to cause grain growth, the grain boundary sweeping of the fission gas atoms from the fuel provides additional gas at the boundary. The grain boundary bubbles grow at a rate determined by the supply of gas atoms at the boundary. A point is reached when the bubbles are large enough to interlink with each other representing grain boundary saturation with fission gases forming a continuous channel for release of the gas to open surfaces. Since the grain boundary saturation is a prerequisite for release of gases in this mechanism, there has to be an incubation period (threshold burn-up) before the gas release occurs. Threshold burn-up for gas release is a function of temperature; higher the temperature lower the burn-up when release starts. Several models of fission gas release (Collins et al 1973; Notley and Hastings 1980) have been developed using the above mechanism.

At temperature above 1600°C almost all the fission gas generated in the fuel is released. The release occurs by movement of gas bubbles which are sufficiently mobile at such high temperatures. The closed pores in the fuel also move up the gradient and sweep the fission gases on their way to the fuel centre where they are released to the central void. The above mechanism of pore and bubble migration for gas release is also responsible for formation of radially oriented columnar grains in the fuel.

In spite of large amount of theoretical and experimental work on fission gas release behaviour in ceramic fuels, the understanding of the mechanism is not complete. There is often a wide gap between the predictions and experimental observations of fission gas release especially in transient tested fuel elements (Pickman 1984). Intensive research in this field, however, unveiled some interesting phenomena in mid-sixties. A scientifically important discovery was that fission events could cause fission gas bubbles to redissolve into the lattice.
Irradiation behaviour of nuclear fuels

Figure 16. The microstructure of an irradiated UO$_2$ fuel from TAPS showing presence of fission gas bubbles on the grain boundary.

4.5 Irradiation creep

Irradiation creep plays a very important role in controlling the strain rate of cladding when fuel-cladding mechanical interaction starts in fuel elements at high burn-up. The effect of irradiation on creep of ceramic fuels is two-fold (a) to enhance the thermal creep, and (b) to develop creep under conditions in which thermal creep is absent. UO$_2$ exhibits creep even at temperatures below 1100°C during irradiation (figure 17). The irradiation induced creep rate is observed to be a linear function of stress and fission rate. Several mechanisms have been proposed for irradiation creep in ceramic fuels (Dienst 1977; Gilbert 1971). The mechanisms proposed for irradiation creep fall in three categories (i) mechanisms based on diffusional creep (ii) mechanisms based on fission spike effect in fuel (iii) mechanisms based on growth of dislocation loops.

It was earlier thought that increase in vacancy concentration resulting from irradiation may enhance the volume self-diffusion term in Nabarro-Herring creep model and this was responsible for the enhancement of creep rate in the fuel. However it was shown (Olander 1976) that creep rate in Nabarro-Herring creep model does not get enhanced because the defects produced by radiation in the grain are either annihilated...
or flow equally to the grain boundary. Brucklacher considered that the point defects formed during irradiation anneal at the defects clusters present inside the grains. He replaced the grain of Nabarro-Herring model by defect clusters of 1000 Å size and was able to predict irradiation creep which was in agreement with the experiment. Another modified diffusional model presented by Brucklacher et al (1970) for irradiation creep considered diffusion of vacancies between dislocation loops formed during irradiation having different orientations relative to the stress direction. Tensile stress causes growth of loops which are normal to the stress, at the expense of loops parallel to the stress direction.

One of the spike-based mechanisms proposed that creep is accelerated by transient pulse heating of fuel by fission spikes. As the material in cylindrical region around the axis of the fission fragment track is momentarily brought to liquid state there is a very large increase in the vacancy diffusion coefficient which gives acceleration of creep. Perrin has proposed another mechanism based on spike effect which is applicable to irradiation enhanced creep in fuel. In this mechanism the fuel undergoes continuous annealing due to fission spikes; strain hardening is prevented and fuel creeps continuously in primary stage (figure 18). The creep rate is given by the primary thermal creep rate and is governed by temperature and fission rate. Creep is considered enhanced with respect to the secondary thermal creep rate which is normally lower than primary creep rate due to strain hardening. Besides the above mechanism the elastic stress relaxation mechanism described in §3.2 has also been proposed for ceramic fuels.

Dislocation-based creep mechanisms consider the formation and growth of dislocation loops under irradiation. It is considered that under the influence of stress
the irradiation-produced defect clusters collapse into dislocation loops preferentially on certain planes. Appropriately oriented loops grow by capture of interstitials. The creep rate in this mechanism is given as a function of stress and fission rate.

5. Prediction of fuel element behaviour

The behaviour of nuclear fuel elements during irradiation is a function of large number of processes which take place simultaneously and interact with each other in a complex manner. Figure 19 shows the complicated interrelation between various physical, chemical, mechanical and metallurgical phenomena which occur inside a water reactor fuel element. To predict the irradiation behaviour of the fuel element a computer model is needed which represents the interactive processes shown in this figure and allows a quantitative estimation of the evolution of changes in the fuel element. The computer model incorporates all the contemporary data of the material properties and the latest understanding of the phenomena and their mutual interactions. It is a very useful tool for design of fuel elements, safety analysis and for the interpretation of the results of irradiation experiments.

The analysis performed by fuel element computer models can be grouped in two categories: thermal analysis and mechanical analysis. The thermal analysis concerns the calculations of the fuel temperature distribution in the fuel set up by the volumetric heat generation. In a cylindrical fuel element, the steady state radial temperature distribution is obtained by solution of following equation with appropriate boundary conditions.

$$\frac{1}{r} \frac{d}{dr} \left( r k \frac{dT}{dr} \right) + H = 0,$$

where $k$ is the thermal conductivity of fuel which is a function of temperature and depends on fuel density. $H$ is volumetric heat generation rate. The heat generation rate in thermal reactor fuel elements is not uniform along the fuel radius owing to thermal
neutron flux depression towards the fuel centre. With the progress of irradiation, the magnitude and radial variation of fuel thermal conductivity \( k \) and heat generation rate \( H \) changes with time due to several irradiation effects e.g. porosity redistribution, fissile material segregation, restructuring etc. The temperature distribution is evaluated at small time intervals accounting for the irradiation effects in the fuel element.

One important aspect of calculation of temperature distribution in fuel is the heat transfer across the fuel cladding gap or interface. The fuel centre temperature is very sensitive to the gap conductance which governs the temperature drop in the fuel-clad gap. The heat transfer coefficient \( h \) for conduction of heat through the gas filled gap is given by

\[
h = \frac{K_{\text{gas}}}{X_{\text{gap}} + g_f + g_t + \delta},
\]

where \( K_{\text{gas}} \) is the thermal conductivity of the gas, \( X_{\text{gap}} \) is the fuel cladding gap in operating conditions, \( \delta \) is the sum of surface roughness of the fuel and the cladding, \( g_f \) and \( g_t \) are jump distance of the gas in fuel and cladding. The value of \( K_{\text{gas}} \) decreases considerably by the release of fission gas to the fuel cladding gap since xenon has a very low thermal conductivity compared to helium which is used as a filler gas in the fuel element. The value of \( X_{\text{gap}} \) during irradiation is controlled by the processes responsible for gap closure e.g. densification, swelling and creep. Uncertainty in the prediction of gas release and fuel clad gap leads to uncertainty in the fuel temperature.
The mechanical analysis is carried out assuming axisymmetry and plane strain conditions. The total strain in the fuel and the cladding is considered to be composed of elastic, thermal, swelling, creep and plastic strains. The constitutive equations used in the modelling are

\[
\varepsilon_r = \frac{1}{E} \left[ \sigma_r - v(\sigma_\theta + \sigma_z) \right] + \varepsilon_r^{th} + \varepsilon_r^e + \varepsilon_r^C + \varepsilon_r^{pl},
\]

\[
\varepsilon_\theta = \frac{1}{E} \left[ \sigma_\theta - v(\sigma_r + \sigma_z) \right] + \varepsilon_\theta^{th} + \varepsilon_\theta^e + \varepsilon_\theta^C + \varepsilon_\theta^{pl},
\]

\[
\varepsilon_z = \frac{1}{E} \left[ \sigma_z - v(\sigma_r + \sigma_\theta) \right] + \varepsilon_z^{th} + \varepsilon_z^e + \varepsilon_z^C + \varepsilon_z^{pl},
\]

where \( v \) is the Poisson’s ratio, \( \varepsilon_r, \varepsilon_\theta, \) and \( \varepsilon_z \) are strains in principal directions \( \sigma_r, \sigma_\theta, \sigma_z \) are stresses in principal directions, \( E \) is elastic modulus, \( \varepsilon^{th} \) is thermal strain, \( \varepsilon^e \) is strain due to swelling, \( \varepsilon^C \) is creep strain and \( \varepsilon^{pl} \) is plastic strain. Swelling and thermal strains are isotropic and hence \( \varepsilon_r^{th} = \varepsilon_\theta^{th} = \varepsilon_z^{th} = \alpha T \) where \( \alpha \) is linear thermal expansions coefficient and

\[
\varepsilon^e_r = \varepsilon^e_\theta = \varepsilon^e_z = \frac{1}{3} \text{ (volumetric swelling)}.
\]

Creep and plastic strains are calculated using equivalent stress \( \sigma^* \) which is defined as

\[
\sigma^* = \frac{1}{\sqrt{2}} \left[ (\sigma_r - \sigma_\theta)^2 + (\sigma_\theta - \sigma_z)^2 + (\sigma_z - \sigma_r)^2 \right]^{1/2}.
\]

Though the constitutive equations remain the same, codes developed may differ in degree of details, numerical methods used and processes modelled.

The fuel modelling codes developed to describe the behaviour of nuclear fuel elements fall in two general groups. In the first category, the models are more phenomenological and consist of a fast-running, empirical models of processes to describe the fuel behaviour. In the second category of codes, the models are treated in more detail and include the various processes which are considered to control the behaviour of the fuel elements. The main objective of such codes is to relate the material properties and physical mechanism to the widest possible range of fuel element geometries, material parameters and operating conditions. Because of the firm physical basis for the models, these codes allow extrapolation over a wide range of conditions.

A number of codes have been developed for nuclear fuel elements all over the world. Some of the well known fuel modelling codes are COMETE (Belgium), ELESIM (Canada), MINIPAT and SLEUTH SEER (UK), URANUS (FRG), LIFE (USA). In India two computer codes have been developed for analysis of water reactor fuel element performance, PROFESS (Sah and Venkatesh 1983, 1984) and COMTA (Anand et al 1980). PROFESS is used for analysis and interpretation of post-irradiation examination results obtained on fuel elements (Roy et al 1984).

6. Summary

A review of the important effects observed in metallic uranium and ceramic nuclear fuels during irradiation has been made. The mechanism of the processes occurring in the fuel during irradiation has been discussed. The basic cause of several effects in fuel
arises from the effects of highly ionised fission fragments and high energy neutrons. The fission fragments give rise to fission spike, lattice defects and impurity atoms in the fuel.

Alpha uranium exhibits dimensional instability during irradiation. This is caused by three irradiation-induced phenomena namely irradiation growth (shape change at constant volume without any external stress application), irradiation creep (deformation under small external stress) and swelling (due to solid and gaseous fission products). Irradiation growth of uranium is related to the anisotropy of alpha uranium crystal and the mechanism of growth involves formation of extra layers of interstitial atoms on certain crystallographic planes.

A very severe thermal and radiation environment (fuel temperatures approaching melting point and thermal gradients in the range 3000–5000°C/cm) initiates a number of interactive processes in the ceramic fuel during irradiation. Several spectacular changes in geometry, dimension, chemistry and microstructure of the fuel are observed. Shrinkage and dissolution of as-fabricated pores by fission fragments leads to fuel densification in the early stages of fuel irradiation. The initial microstructure of the fuel gets altered and several restructured radial zones are observed in fuel cross-sections.

One of the important microstructural changes is the formation of high density columnar grain by movement of porosity towards the centre. Plutonium which is initially distributed uniformly in the fuel is observed to segregate to preferred locations (towards fuel centre or away from fuel centre) depending upon the initial stoichiometry of the fuel. Oxygen in the fuel gets redistributed to maintain the thermodynamic equilibrium between CO₂, CO and oxygen in the existing radial temperature gradient. Many fission products especially the noble metals and volatile elements produced in the fuel are also observed to redistribute themselves. A large fraction of fission products in fuel is the fission gases xenon and krypton which are insoluble in the fuel matrix. The fission gases precipitate as gas bubbles in the fuel and cause fuel swelling. A part of the fission gas is released from the fuel. The gas release occurs by three mechanisms: recoil and knock out (below 1000°C); atomic diffusion to the grain boundary (1000–1600°C) and bubble migration to the fuel centre (above 1600°C). Another important phenomena observed in ceramic fuel is irradiation creep. Irradiation-induced creep is found to take place in UO₂ at a low temperature (< 1100°C) where thermal creep is absent. Three types of mechanisms have been proposed: diffusion based, spike based and dislocation based. A satisfactory model of irradiation creep in ceramic fuel is yet to be developed.

References

Assman H and Stehle H 1978 Nucl. Eng. Design 48 49
Bahl J K, Sah D N, Chatterjee S and Sivaramakrishnan K S 1979 BARC Report, BARC-1014
Bellamy R G and Rich J R 1969 J. Nucl. Mater. 33 64
Bramman J I and Powell H J 1975 J. Br. Nucl. Soc. 1 63