NOBEL LECTURES

PHYSICS

1963 – 1970

World Scientific

Published for the Nobel Foundation in 1998 by

World Scientific Publishing Co. Pte. Ltd.
P O Box 128, Farrer Road, Singapore 912805
USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

NOBEL LECTURES IN PHYSICS (1963–1970)

All rights reserved.

ISBN 981-02-3404-X

Printed in Singapore.

Foreword

Since 1901 the Nobel Foundation has published annually *Les Prix Nobel* with reports from the Nobel award ceremonies in Stockholm and Oslo as well as the biographies and Nobel lectures of the Laureates. In order to make the lectures available to people with special interests in the different prize fields the Foundation gave Elsevier Publishing Company the right to publish in English the lectures for 1901–1970, which were published in 1964–1972 through the following volumes:

Physics 1901–1970	4 vols.
Chemistry 1901–1970	4 vols.
Physiology or Medicine 1901–1970	4 vols.
Literature 1901–1967	1 vol.
Peace 1901–1970	3 vols.

Since the Elsevier series has been out of print for many years the Nobel Foundation has given World Scientific Publishing Company the right to publish these Nobel lectures, biographies and presentation speeches. The Nobel Foundation is very pleased that the intellectual and spiritual message to the world laid down in the laureates' lectures will, thanks to the efforts of World Scientific, reach new readers all over the world.

Bengt Samuelsson
Chairman of the Board

Michael Sohlman
Executive Director

Stockholm, March 1998

Contents

	Foreword	v
	Publisher's Note	vii
1963	EUGENE P. WIGNER, MARIA GOEPPERT MAYER and J. HANS D. JENSEN	
	Motivation	1
	Presentation by I. Waller	3
	Events, laws of nature, and invariance principles, by Eugene P. Wigner	6
	Biography Eugene P. Wigner	18
	The shell model, by Maria Goeppert Mayer	20
	Biography Maria Goeppert Mayer	38
	Glimpses at the history of the nuclear structure theory, by J. Hans D. Jensen	40
	Biography J. Hans D. Jensen	51
1964	CHARLES H. TOWNES, NIKOLAI G. BASOV, and ALEXANDER M. PROCHOROV	
	Motivation	53
	Presentation by B. Edlén	55
	Production of coherent radiation by atoms and molecules, by Charles H. Townes	58
	Biography Charles H. Townes	87
	Semiconductor lasers, by Nikolai G. Basov	89
	Biography Nikolai G. Basov	106
	Quantum electronics, by A. M. Prochorov	110
	Biography Alexander M. Prochorov	117

1965 SIN-ITIRO TOMONAGA, JULIAN SCHWINGER, and RICHARD P. FEYNMAN
Motivation 121
Presentation by I. Waller 123
Development of quantum electrodynamics, by Sin-itiro Tomonaga 126
Biography Sin-itiro Tomonaga 137
Relativistic quantum field theory, by Julian Schwinger 140
Biography Julian Schwinger 153
The development of the space-time view of quantum electrodynamics, by Richard P. Feynman 155
Biography Richard P. Feynman 179

1966 ALFRED KASTLER
Motivation 181
Presentation by I. Waller 183
Optical methods for studying Hertzian resonances, by Alfred Kastler 186
Biography Alfred Kastler 205

1967 HANS ALBRECHT BETHE
Motivation 209
Presentation by O. Klein 211
Energy production in stars, by H. A. Bethe 215
Biography Hans Albrecht Bethe 235

1968 LUIS W. ALVAREZ
Motivation 237
Presentation by S. von Friesen 239
Recent developments in particle physics, by Luis W. Alvarez 241
Biography Luis W. Alvarez 291

1969 MURRAY GELL-MANN
Motivation 293
Presentation by I. Waller 295
Symmetry and currents in particle physics, by Murray Gell-Mann 299
Biography Murray Gell-Mann 300

1970 HANNES ALFVÉN and
LOUIS NÉEL
Motivation 301
Presentation by T. Gustafsson 303
Plasma physics, space research and the origin of the solar system by Hannes Alfvén 306
Biography Hannes Alfvén 317
Magnetism and the local molecular field by Louis Néel 318
Biography Louis Néel 342

Name Index 345
Subject Index 350
Index of Biographies 351

Physics 1963

EUGENE P. WIGNER

«for his contributions to the theory of the atomic nucleus and the elementary particles, particularly through the discovery and application of fundamental symmetry principles»

MARIA GOEPPERT MAYER

J. HANS D. JENSEN

«for their discoveries concerning nuclear shell structure»

Physics 1963

Presentation Speech by Professor I. Waller, member of the Nobel Committee for Physics of the Royal Swedish Academy of Sciences

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

The discoveries by Eugene Wigner, Maria Goeppert Mayer and Hans Jensen for which this year's Nobel prize in physics has been awarded, concern the theory of the atomic nuclei and the elementary particles. They are based on the highly successful atomic research of the first three decades of this century which showed that an atom consists of a small nucleus and a surrounding cloud of electrons which revolve around the nucleus and thereby follow laws which had been formulated in the so-called quantum mechanics. To the exploration of the atomic nuclei was given a firm foundation in the early 1930's when it was found that the nuclei are built up by protons and neutrons and that the motion of these so-called nucleons is governed by the laws of quantum mechanics.

In order to be able to calculate the motion of the nucleons it was, however, necessary to know also the forces which act between them. A very important step in the investigation of these forces was taken by Wigner in 1933 when he found, deducing from some experiments, that the force between two nucleons is very weak except when their distance apart is very small but that the force is then a million times stronger than the electric forces between the electrons in the outer part of the atoms. Wigner discovered later other important properties of the nuclear forces.

Nothwithstanding the efforts of many physicists our knowledge of the nuclear forces is yet rather incomplete. It was therefore fundamentally important that Wigner could show that most essential properties of the nuclei follow from generally valid symmetries of the laws of motion. Earlier Wigner had performed pioneering work by studying such symmetries in the laws of motion for the electrons and had made important discoveries by investigating *e.g.* those symmetries which express the fact that the laws mentioned make no difference between left and right and that backward in time according to them is equivalent to forward in time. These investigations were extended by Wigner to the atomic nuclei at the end of the 1930's and he explored then also the newly discovered symmetry property of the force between two nu-

cleons to be the same whether either of the nucleons is a proton or a neutron.

This work by Wigner and his other investigations of the symmetry principles in physics are important far beyond nuclear physics proper. His methods and results have become an indispensable guide for the interpretation of the rich and complicated picture which has emerged from recent years' experimental research on elementary particles. They were also an important preliminary for the deeper penetration into and the partial revision of the earlier concepts concerning the right-left symmetry which was accomplished by Yang and Lee. They were therefore awarded the Nobel Prize in Physics of 1957.

Wigner has made many other important contributions to nuclear physics. He has given a general theory of nuclear reactions and has made decisive contributions to the practical use of nuclear energy. He has, often in collaboration with younger scientists, broken new paths in many other domains of physics.

An initially independent line of research in nuclear physics has been the attempts to find models for the atomic nuclei which visualize the motion of the nucleons.

It was found during the 1920's and in particular during the 1930's that the protons and the neutrons each form particularly stable systems in an atomic nucleus when the numbers of either kind of nucleons is one of the so-called magic numbers 2, 8, 20, 50, 82 and 126. Several physicists, in particular Elsasser, tried to interpret the magic numbers in analogy to Niels Bohr's successful explanation of the periodic system of the elements. It was then assumed that the nucleons move in orbits in a common field of force and that these orbits are arranged in so-called shells which are energetically well separated. The magic members should correspond to complete shells. This interpretation was successful for light nuclei. It was, however, not possible to explain more than the three first magic numbers and for many years another model dominated.

A paper published by Goeppert Mayer in 1948 marked the beginning of a new era in the appreciation of the shell model. For the first time convincing evidence was there given for the existence of the higher magic numbers and it was stressed that the experiments support the existence of closed shells very strongly.

Somewhat later Goeppert Mayer and independently Haxel, Jensen and Suess published the new idea, which was needed for the explanation of the higher magic numbers. The idea was that a nucleon should have different energies when it «spins» in the same or opposite sense as it revolves around the nucleus.

Goeppert Mayer and Jensen collaborated later on the development of the shell model. They published together a book, where they applied the model to the extensive experimental material on atomic nuclei. They gave convincing evidence for the great importance of the shell model in systematizing this material and predicting new phenomena concerning the ground state and the low excited states of the nuclei. The general methods introduced by Wigner have been most important for the applications of the shell model. It has also been possible to give a deeper justification of the shell model. Its fundamental importance has thereby been further confirmed.

Professor Wigner. In the late 1920's you laid the foundation of the theory of symmetries in quantum mechanics and introduced new ideas and methods which you have later further developed and successfully applied to some of the most fundamental problems in physics. This work and your other contributions have been of farreaching importance, indeed essential, for the development of the nuclear and elementary particle physics of our time.

Professor Goeppert Mayer, Professor Jensen. Your work on the shell model which you started independently and then pursued in collaboration has shed new light on the structure of atomic nuclei. It constitutes a most striking advance in the correlation of nuclear properties. Your work has inspired an ever increasing number of new investigations and has been indispensable for the later work, both experimental and theoretical, on atomic nuclei.

Professor Wigner, Professor Goeppert Mayer, Professor Jensen. On behalf of the Academy I wish to extend to you our hearty congratulations and now ask you to receive from the Hands of His Majesty the King the Nobel Prize for Physics for the year 1963.

EUGENE P. WIGNER

Events, laws of nature, and invariance principles

Nobel Lecture, December 12, 1963

It is a great and unexpected honor to have the opportunity to speak here today. Six years ago, Yang and Lee spoke here, reviewing symmetry principles in general and their discovery of the violation of the parity principle in particular[1]. There is little point in repeating what they said, on the history of the invariance principles, or on my own contribution to these which they, naturally, exaggerated. What I would like to discuss instead is the general role of symmetry and invariance principles in physics, both modern and classical. More precisely, I would like to discuss the relation between three categories which play a fundamental role in all natural sciences: events, which are the raw materials for the second category, the laws of nature, and symmetry principles for which I would like to support the thesis that the laws of nature form the raw material.

Events and Laws of Nature

It is often said that the objective of physics is the explanation of nature, or at least of inanimate nature. What do we mean by explanation? It is the establishment of a few simple principles which describe the properties of what is to be explained. If we understand something, its behavior, that is the events which it presents, should not produce any surprises for us. We should always have the impression that it could not be otherwise.

It is clear that, in this sense, physics does not endeavor to explain nature. In fact, the great success of physics is due to a restriction of its objectives: it only endeavors to explain the *regularities* in the behavior of objects. This renunciation of the broader aim, and the specification of the domain for which an explanation can be sought, now appears to us an obvious necessity. In fact, the specification of the explainable may have been the greatest discovery of physics so far. It does not seem easy to find its inventor, or to give the exact date of its origin. Kepler still tried to find exact rules for the magnitude of the planetary orbits, similar to his laws of planetary motion. Newton already

realized that physics would deal, for a long time, only with the explanation of those of the regularities discovered by Kepler which we now call Kepler's laws[2].

The regularities in the phenomena which physical science endeavors to uncover are called the laws of nature. The name is actually very appropriate. Just as legal laws regulate actions and behavior under certain conditions, but do not try to regulate all actions and behavior, the laws of physics also determine the behavior of its objects of interest only under certain well defined conditions, but leave much freedom otherwise. The elements of the behavior which are not specified by the laws of nature are called initial conditions. These, then, together with the laws of nature, specify the behavior as far as it can be specified at all: if a further specification were possible, this specification would be considered as an added initial condition. As is well known, before the advent of quantum theory, it was believed that a complete description of the behavior of an object is possible so that, if classical theory were valid, the initial conditions and the laws of nature together would completely determine the behavior of an object.

The preceding statement is a definition of the term «initial condition». Because of its somewhat unusual nature, it may be worthwhile to illustrate this on an example. Suppose we did not know Newton's equation for the motion of stars and planets

$$\ddot{\mathfrak{r}}_i = G\,\Sigma' M_j\,\frac{\mathfrak{r}_{ij}}{r_{ij}^3} \qquad \mathfrak{r}_{ij} = \mathfrak{r}_j - \mathfrak{r}_i \tag{1}$$

but had found only the equation determining the third derivative of the position

$$\dddot{\mathfrak{r}}_i = G\Sigma' M_j\,\frac{\dot{\mathfrak{r}}_{ij}(\mathfrak{r}_{ij}\cdot\mathfrak{r}_{ij}) - 3\,\mathfrak{r}_{ij}(\dot{\mathfrak{r}}_{ij}\cdot\mathfrak{r}_{ij})}{r_{ij}^5} \tag{2}$$

More generally, if the forces F_i are non-gravitational, one would have written

$$M_i\,\ddot{\mathfrak{r}}_i = \dot{\mathfrak{r}}_i\cdot \mathrm{grad} F_i + \dot{F}_i \tag{2a}$$

The initial conditions then would contain not only all the $\mathfrak{r}_i$ and $\dot{\mathfrak{r}}_i$, but also the $\ddot{\mathfrak{r}}_i$. These data, together with the «equation of motion» (2), would then determine the future behavior of the system just as r_i, $\dot{r}_i$ and (1) determines it. The fact that initial conditions and laws of nature completely determine the behavior is, similarly, true in any causal theory.

The surprising discovery of Newton's age is just the clear separation of laws

of nature on the one hand and initial conditions on the other. The former are precise beyond anything reasonable; we know virtually nothing about the latter. Let us pause for a minute at this last statement. Are there really no regularities concerning what we just called initial conditions?

The last statement would certainly not be true if the laws of nature (2), (2a) were adopted, *i.e.*, if we considered the $\ddot{r}_i$ as part of the initial conditions. In this case, there would be a relation, in fact the precise relation (1), between the elements of the initial conditions. The question, therefore, can be only: are there any relations between what we really do consider as initial conditions. Formulated in a more constructive way: how can we ascertain that we know all the laws of nature relevant to a set of phenomena? If we do not, we would determine unnecessarily many initial conditions in order to specify the behavior of the object. One way to ascertain this would be to prove that all the initial conditions can be chosen arbitrarily – a procedure which is, however, impossible in the domain of the very large (we cannot change the orbits of the planets), or the very small (we cannot precisely control atomic particles). No other equally unambiguous criterion is known to me, but there is a distinguishing property of the correctly chosen, that is minimal set, of initial conditions which is worth mentioning.

The minimal set of initial conditions not only does not permit any exact relation between its elements, on the contrary, there is reason to contend that these are, as a rule, as random as the externally imposed, gross constraints allow. I wish to illustrate this point first on an example which, at first, seems to contradict it because this example shows its power, and also its weakness, best.

Let us consider for this purpose again our planetary system. It was mentioned before that the approximate regularities in the initial conditions, that is the determinants of the orbits, led Kepler to the considerations which were then left by the wayside by Newton. These regularities form the apparent counterexample to the aforementioned thesis. However, the existence of the regularities in the initial conditions is considered so unsatisfactory that it is felt necessary to show that the regularities are but a consequence of a situation in which there were *no* regularities. Perhaps v. Weizsäcker's attempt in this direction[3] is most interesting: he assumes that, originally, the solar system consisted of a central star, with a gas in rotation, but otherwise in random motion, around it. He then deduces the aforementioned regularities of the planetary system, now called Bode's law, from his assumption. More generally, one tries to deduce almost all «organized motion», even the existence of life, in a similar

fashion. It must be admitted that few of these explanations have been carried out in detail[4] but the fact that such explanations are attempted remains significant.

The preceding paragraph dealt with cases in which there is at least an apparent evidence against the random nature of the uncontrolled initial conditions. It attempted to show that the apparently organized nature of these initial conditions was preceded by a state in which the uncontrolled initial conditions were random. These are, on the whole, exceptional situations. In most cases, there is no reason to question the random nature of the non-controlled, or non-specified, initial conditions and the random nature of these initial conditions is supported by the validity of the conclusions arrived at on the basis of the assumption of randomness. One encounters such situations in the kinetic theory of gases and, more generally, whenever one describes processes in which the entropy increases. Altogether, then, one obtains the impression that whereas the laws of nature codify beautifully simple regularities, the initial conditions exhibit, as far as they are not controlled, equally simple and beautiful irregularity. Hence, there is perhaps little chance that some of the former remain overlooked.

The preceding discussion characterized the laws of nature as regularities in the behavior of an object. In quantum theory, this is natural: the laws of quantum mechanics can be suitably formulated as correlations between subsequent observations on an object. These correlations are the regularities given by the laws of quantum mechanics[5]. The statement of classical theory, its equations of motion, are not customarily viewed as correlations between observations. It is true, however, that their purpose and function is to furnish such correlations and that they are, in essence, nothing but a shorthand expression for such correlations.

Laws of Nature and Invariance

We have ceased to expect from physics an explanation of all events, even of the gross structure of the universe, and we aim only at the discovery of the laws of nature, that is the regularities, of the events. The preceding section gives reason for the hope that the regularities form a sharply defined set, and are clearly separable from what we call initial conditions, in which there is a strong element of randomness. However, we are far from having found that set. In fact, if it is true that there are precise regularities, we have reason to

believe that we know only an infinitesimal fraction of these. The best evidence for this statement derives perhaps from a fact which was mentioned here by Yang six years ago: the multiplicity of the types of interactions. Yang mentioned four of them: gravitational, weak, electromagnetic, and strong, and it now seems that there are two types of strong interactions. All these play a role in every process, but it is hard, if not impossible, to believe that the laws of nature should have such complexity as implied by four or five different types of interactions between which no connection, no analogy, can be discovered.

It is natural, therefore, to ask for a superprinciple which is in a similar relation to the laws of nature as these are to the events. The laws of nature permit us to foresee events on the basis of the knowledge of other events; the principles of invariance should permit us to establish new correlations between events, on the basis of the knowledge of established correlations between events. This is exactly what they do. If it is established that the existence of the events A, B, C,... necessarily entails the occurrence of X, then the occurrence of the events A′, B′, C′,... also necessarily entails X′, if A′, B′, C′,... and X′ are obtained from A, B, C,... and X by one of the invariance transformations. There are three categories of such invariance transformations:

(a) euclidean transformations: the primed events occur at a different location in space, but in the same relation to each other, as the unprimed events.

(b) time displacements: the primed events occur at a different time, but separated by the same time intervals from each other, as the unprimed ones.

(c) uniform motion: the primed events appear to be the same as the unprimed events from the point of view of a uniformly moving coordinate system.

The first two categories of invariance principles were always taken for granted. In fact, it may be argued that laws of nature could not have been recognized if they did not satisfy some elementary invariance principles such as those of Categories (a) and (b)–if they changed from place to place, or if they were also different at different times. The principle (c) is not so natural. In fact, it has often been questioned and it was an accomplishment of extraordinary magnitude, on the part of Einstein, to have reestablished it in his special theory of relativity. However, before discussing this point further, it may be useful to make a few general remarks.

The first remarkable characteristic of the invariance principles which were enumerated is that they are all geometric, at least if four-dimensional space-time is the underlying geometrical space. By this I mean that the invariance transformations do not change the events; they only change their location in space and time, and their state of motion. One could easily imagine a prin-

ciple in which, let us say, protons are replaced by electrons and *vice versa*, velocities by positions, and so on[6].

The second remarkable characteristic of the preceding principles is that they are invariance rather than covariance principles. This means that they postulate the same conclusion for the primed premises as for the unprimed premises. It is quite conceivable that, if certain events A, B,... take place, the events X_1, X_2, X_3... will follow with certain probabilities p_1, p_2, p_3... From the transformed events A′, B′, C′, the transformed consequences X_1', X_2', X_3'... could follow with changed probabilities such as $p_1' = p_1(1 - p_1 + \Sigma p_n^2)$, $p_2' = p_2(1 - p_2 + \Sigma p_n^2)$,... but this is not the case; we always had $p_i' = p_i$.

These two points are specifically mentioned because there are symmetry principles, the so-called crossing relations[7], which *may be* precisely valid and which surely do not depend on specific types of interactions. In these regards they are, or may be, similar to the geometric invariance principles. They differ from these because they do change the events and they are covariance rather than invariance principles. Thus, from a full knowledge of the cross section for neutron–proton scattering, they permit one to obtain some of the neutron–antineutron collision cross sections. The former events are surely different from the neutron–antineutron collisions and the cross sections for the latter are not equal to the neutron–proton cross sections but are obtained from these by a rather complicated mathematical procedure. Hence, the crossing relations are not considered to be geometrical symmetry conditions and they will not be considered here. Similarly, we shall not be concerned with the dynamic symmetry principles which are symmetries of specific interactions, such as electromagnetic interactions or strong interactions, and are not formulated in terms of events[7].

On the other hand, it should be noted that the invariance principles themselves depend on the dividing line between initial conditions and laws of nature. Thus, the law of nature (2) or (2a), obtained from Newton's principle by differentiation with respect to time, is invariant also under the transformation to a uniformly accelerated coordinate system

$$\mathfrak{r}_i' = \mathfrak{r}_i + t^2 a \qquad t' = t \tag{3}$$

where a is an arbitrary vector. Naturally, this added principle can have no physical consequence because, if the initial conditions $\mathfrak{r}_i$, $\dot{\mathfrak{r}}_i$, $\ddot{\mathfrak{r}}_i$ are realizable (*i.e.*, satisfy (1)), the transformed initial conditions $\mathfrak{r}_i' = \mathfrak{r}_i$, $\dot{\mathfrak{r}}_i' = \dot{\mathfrak{r}}_i$, $\ddot{\mathfrak{r}}_i' = \ddot{\mathfrak{r}}_i + 2a$ cannot be realizable.

The symmetry principles of the preceding discussion are those of New-

tonian mechanics or the special theory of relativity. One may well wonder why the much more general, and apparently geometrical, principles of invariance of the general theory have not been discussed. The reason is that this writer believes, in conformity with the views expressed by V. Fock[8], that the curvilinear coordinate transformations of the general theory of relativity are not invariance transformations in the sense considered here. These were so-called active transformations, replacing events A, B, C,... by events A′, B′, C′,... and unless active transformations are possible, there is no physically meaningful invariance. However, the mere replacement of one curvilinear coordinate system by another is a «redescription» in the sense of Melvin[9]; it does not change the events and does not represent a structure in the laws of nature. This does not mean that the transformations of the general theory of relativity are not useful tools for finding the correct laws of gravitation; they evidently are. However, as I suggested elsewhere[7], the principle which they serve to formulate is different from the geometrical invariance principles considered here.

The Use of Invariance Principles, Approximate Invariances

The preceding two sections emphasized the inherent nature of the invariance principles as being rigorous correlations between those correlations between events which are postulated by the laws of nature. This at once points to the use of the set of invariance principles which is surely most important at present: to be a touchstone for the validity of possible laws of nature. A law of nature can be accepted as valid only if the correlations which it postulates are consistent with the accepted invariance principles.

Incidentally, Einstein's original article which led to his formulation of the special theory of relativity illustrates the preceding point with greatest clarity[10]. He points out in this article that the correlations between events are the same in coordinate systems in uniform motion with respect to each other, even though the causes *attributed* to these correlations at that time did depend on the state of motion of the coordinate system. Similarly, Einstein made the most extensive use of invariance principles to guess the correct form of a law of nature, in this case that of the gravitational law, by postulating that this law conform with the invariance principles which he postulated[11]. Equally remarkable is the present application of invariance principles in quantum electrodynamics. This is not a consistent theory – in fact, not a theory in the proper sense because its equations are in contradiction to each other. However, these

contradictions can be resolved with reasonable uniqueness by postulating that the conclusions conform to the theory of relativity[12]. Another approach, even more fundamental, tries to axiomatize quantum field theories, the invariance principles forming the cornerstone of the axioms[13]. I will not further enlarge on this question because it has been discussed often and eloquently. In fact, I myself spoke about it but a short time ago[7].

To be touchstones for the laws of nature is probably the most important function of invariance principles. It is not the only one. In many cases, consequences of the laws of nature can be derived from the character of the mathematical framework of the theory, together with the postulate that the law – the exact form of which need not be known – conform with invariance principles. The best known example herefor is the derivation of the conservation laws for linear and angular momentum, and for energy, and the motion of the center of mass, either on the basis of the Lagrangian framework of classical mechanics, or the Hilbert space of quantum mechanics, by means of the geometrical invariance principles enumerated before[14]. Incidentally, conservation laws furnish at present the only generally valid correlations between observations with which we are familiar; for those which derive from the geometrical principles of invariance it is clear that their validity transcends that of any special theory – gravitational, electromagnetic, etc. – which are only loosely connected in present-day physics. Again, the connection between invariance principles and conservation laws – which in this context always include the law of the motion of the center of mass – has been discussed in the literature frequently and adequately.

In quantum theory, invariance principles permit even further reaching conclusions than in classical mechanics and, as a matter of fact, my original interest in invariance principles was due to this very fact. The reason for the increased effectiveness of invariance principles in quantum theory is due, essentially, to the linear nature of the underlying Hilbert space[15]. As a result, from any two state vectors, Ψ_1 and Ψ_2, an infinity of new state vectors

$$\Psi = a_1\Psi_1 + a_2\Psi_2 \tag{4}$$

can be formed, a_1 and a_2 being arbitrary numbers. Similarly, several, even infinitely many, states can be superimposed with arbitrary coefficients. This possibility of superposing states is by no means natural physically. In particular even if we know how to bring a system into the states Ψ_1 and Ψ_2, we cannot give a prescription how to bring it into a superposition of these states. This prescription would have to depend, naturally, on the coefficients with which

the two states are superimposed and is simply unknown. Hence, the superposition principle is strictly an existence postulate–but a very effective and useful existence postulate.

To illustrate this point, let us note that in classical theory, if a state, such as a planetary orbit, is given, another state, that is another orbit, can be produced by rotating the initial orbit around the center of attraction. This is interesting but has no very surprising consequences. In quantum theory, the same is true. In addition, however, the states obtained from a given one by rotation can be superimposed as a result of the aforementioned principle. If the rotations to which the original state was subjected are uniformly distributed over all directions, and if the states so resulting are superimposed with equal coefficients, the resulting state has necessarily spherical symmetry. This is illustrated in the Fig. 1 in the plane case. This construction of a spherically symmetric state could fail only if the superposition resulted in the null-vector of Hilbert space in which case one would not obtain any state. In such a case, however, other coefficients could be chosen for the superposition–in the plane case the coefficients $e^{im\varphi}$ where φ is the angle of rotation of the original state–and the resulting state, though not spherically symmetric, or in the plane case axially symmetric – would still exhibit simple properties with respect to rotation. This possibility, the construction of states which have either full rotational symmetry, or at least some simple behavior with respect to rotations, is the one which is fundamentally new in quantum theory. It is also conceptually satisfying that simple systems, such as atoms, have states of high symmetry.

The superposition principle also permits the exploitation of reflection sym-

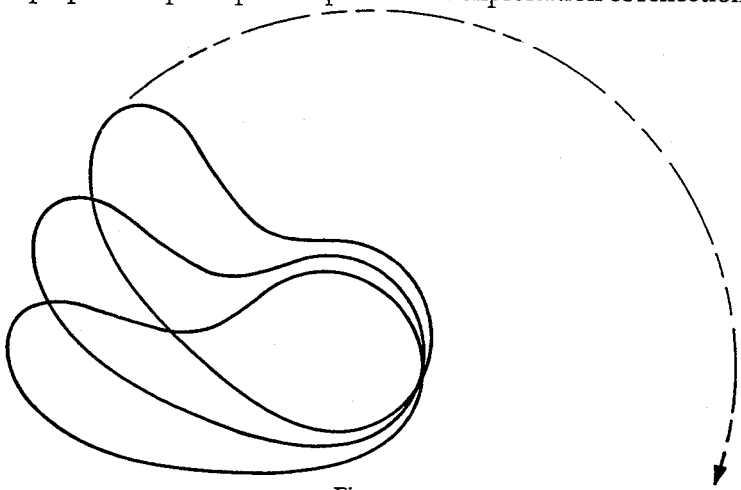


Fig. 1.

metry. In classical mechanics as well as in quantum mechanics, if a state is possible, the mirror image of that state is also possible. However, in classical theory no significant conclusion from this fact is possible. In quantum theory, original state and mirror image can be superimposed, with equal or oppositely equal coefficients. In the first case, the resulting state is symmetric with respect to reflections, in the second case antisymmetric. The great accomplishment of Lee and Yang, which was mentioned earlier[1], was just a very surprising reinterpretation of the physical nature of one of the reflection operations, that of space reflection, with the additional proof that the old interpretation cannot be valid. The consideration of «time inversion» requires rather special care because the corresponding operator is antiunitary. Theoretically, it does lead to a new quantum number and a classification of particles[16] which, however, has not been applied in practice.

My discussion would be far from complete without some reference to approximate invariance relations. As all approximate relations, these may be very accurate under certain conditions but fail significantly in others. The critical conditions may apply to the state of the object, or may specify a type of phenomena. The most important example for the first case is that of low velocities. In this case, the magnetic fields are weak and the direction of the spins does not influence the behavior of the other coordinates. One is led to the Russel-Saunders coupling of spectroscopy[17]. Even more interesting should be the case of very high velocities in which the magnitude of the rest mass becomes unimportant. Unfortunately, this case has not been discussed in full detail even though there are promising beginnings[18].

Perhaps the most important case of special phenomena in which there are more invariance transformations than enumerated before is rather general: it comprises all phenomena, such as collisions between atoms, molecules, and nuclei, in which the weak interaction, which is responsible for beta decay, does not play a role. In all these cases, the parity operation is a valid invariance operation. This applies also in ordinary spectroscopy.

In another interesting special type of phenomena the electromagnetic interaction also plays a subordinate role only. This renders the electric charge on the particles insignificant and the interchange of proton and neutron, or more generally of the members of an isotopic spin multiplet, becomes an invariance operation. These, and the other special cases of increased symmetry, lead to highly interesting questions which are, furthermore, at the center of interest at present. However, the subject has too many ramifications to be discussed in detail at this occasion.

1. Chen Ning Yang, The law of parity conservation and other symmetry laws of physics; Tsung Dao Lee, Weak interactions and nonconservation of parity, *Nobel Lectures, Physics 1942–1962*, Elsevier, Amsterdam, 1964, pp. 393–403 and pp. 406–418.
2. See, for instance, A. C. Crombie, *Augustine to Galileo*, Falcon Press, London, 1952, pp. 316 ff. The growth of the understanding of the realm of the explainable, from the end of the 13th century on, can be traced through almost every chapter of this book.
3. C. F. v. Weizsäcker, *Z. Astrophys.*, 22 (1944) 319; S. Chandrasekhar, *Rev. Mod. Phys.*, 18 (1946) 94.
4. An interesting and well understood case is that of «focussing collisions» in which neutrons, having velocities which are rather high but with random orientation, are converted into lower velocity neutrons but with preferential directions of motion. See R. H. Silsbee, *J. Appl. Phys.*, 28 (1957) 1246; Chr. Lehmann and G. Leibfried, *Z. Physik*, 172 (1963) 465.
5. See, for instance, the section *What is the State Vector* in the writer's article, *Am. J. Phys.*, 31 (1963) 6.
6. The possibility of an invariance principle in which velocities are replaced by positions, and conversely, was studied by M. Born, *Nature*, 141 (1938) 327; *Proc. Roy. Soc. (London), Ser. A*, 165 (1938) 291, 166 (1938) 552.
7. The crossing relations were established by M. L. Goldberger, *Phys. Rev.*, 99 (1955) 979; M. Gell-Mann and M. L. Goldberger, *Phys. Rev.*, 96 (1954) 1433. For further literature, see, for instance M. L. Goldberger and K. M. Watson, *Collision Theory*, Wiley, New York, 1964, chapter 10. The relations of the various types of symmetry principles were considered in two recent articles: *Nuovo Cimento, Suppl.*, in the press, and *Phys. Today*, 17 (1964) 34. See also E. Wigner, *Progr. Theoret. Phys. (Kyoto)*, 11 (1954) 437.
8. V. A. Fock, *The Theory of Space, Time and Gravitation*, Pergamon, Oxford, 1959. The character of the postulate of invariance with respect to general coordinate transformations as a geometrical invariance was questioned already by E. Kretschman, *Ann. Physik*, 53 (1917) 575.
9. M. A. Melvin, *Rev. Mod. Phys.*, 32 (1960) 477.
10. A. Einstein, Zur Elektrodynamik bewegter Körper, *Ann. Physik*, 17 (1905) 891.
11. A. Einstein and S. B. Preuss, *Akad. Wiss.*, (1915) 778, 779, 844; *Ann. Physik*, 49 (1916) 769. Similar results were obtained almost simultaneously by D. Hilbert, *Nachr. kgl. Ges. Wiss. (Göttingen)*, (1915) 395.
12. J. Schwinger, *Phys. Rev.*, 76 (1949) 790. See also S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory*, Row, Peterson and Co., New York, 1961, Section 15, where further references can also be found.
13. See A. S. Wightman, Quelques Problèmes Mathématiques de la Théorie Quantique Relativiste, and numerous other articles in *Les Problèmes Mathématiques de la Théorie Quantique des Champs*, Centre National de la Recherche Scientifique, Paris, 1959.
14. G. Hamel, *Z. Math. Physik*, 50 (1904) 1; G. Herzglotz, *Ann. Physik*, 36 (1911) 493; F. Engel, *Nachr. kgl. Ges. Wiss. (Göttingen)*, (1918) 171; E. Noether, *Nachr. kgl. Ges. Wiss. (Göttingen)*, (1918) 235; E. Bessel-Hagen, *Math. Ann.*, 84 (1921) 258; the quantum theoretical derivation given by E. Wigner, *Nachr. kgl. Ges. Wiss. (Göttingen)*,

(1927) 375, contains also the parity conservation law which was shown, in ref.1, to be only approximately valid. See also the article of ref.16.

15. I heard this remark, for the first time, from C. N. Yang, at the centennial celebration of Bryn Mawr College.
16. See the writer's article Unitary Representations of the Inhomogeneous Lorentz Group including Reflections, in *Elementary Particle Physics*, Gordon and Breach, New York, 1964, for a systematic discussion of the reflection operations.
17. See the writer's book, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*, Vieweg, Braunschweig, 1931, or the English translation by J. Griffin, Academic Press, New York, 1959.
18. H. A. Kastrup, *Physics Letters*, 3 (1962) 78. The additional invariance operations probably form the conformal group. This was discovered by E. Cunningham (*Proc. London Math. Soc.*, 8 (1909) 77) and by H. Bateman (*Proc. London Math. Soc.*, 8 (1910) 223) to leave Maxwell's equations for the vacuum invariant, *i.e.*, the equations which describe light, always propagating at light velocity. For more recent considerations, see T. Fulton, F. Rohrlich and L. Witten, *Rev. Mod. Phys.*, 34 (1962) 442 and Y. Murai, *Progr. Theoret. Phys.* (*Kyoto*), 11 (1954) 441. The latter articles contain also more extensive references to the subject.

Biography

Eugene Paul Wigner, born in Budapest, Hungary, on November 17, 1902, naturalized a citizen of the United States on January 8, 1937, has been since 1938 Thomas D. Jones Professor of Mathematical Physics at Princeton University–he retired in 1971. His formal education was acquired in Europe; he obtained the Dr. Ing. degree at the Technische Hochschule Berlin. Married in 1941 to Mary Annette Wheeler, he is the father of two children, David and Martha. His son, David, is teaching mathematics at the University of California in Berkeley. His daughter, Martha, is with the Chicago area transportation system, an organization endeavoring to improve the internal transportation system of that city. Dr. Wigner worked on the Manhattan Project at the University of Chicago during World War II, from 1942 to 1945, and in 1946–1947 became Director of Research and Development at Clinton Laboratories. Official recognition of his work in nuclear research includes the U. S. Medal for Merit, presented in 1946; the Enrico Fermi Prize (U. S. A. E. C.) awarded in 1958; and the Atoms for Peace Award, in 1960. Dr. Wigner holds the Medal of the Franklin Society, the Max Planck Medal of the German Physical Society, the George Washington Award of the American–Hungarian Studies Foundation (1964), the Semmelweiss Medal of the American–Hungarian Medical Association (1965), and the National Medal of Science (1969). He has received honorary degrees from the University of Wisconsin, Washington University, Case Institute, University of Alberta (Canada), University of Chicago, Colby College, University of Pennsylvania, Yeshiva University, Thiel College, Notre Dame University, Technische Universität Berlin, Swarthmore College, Université de Louvain, Université de Liège, University of Illinois, Seton Hall, Catholic University and The Rockefeller University. He is a past vice-president and president of the American Physical Society, of which he remains a member. He is a past member of the board of directors of the American Nuclear Society and still a member; he holds memberships in the American Philosophical Society, the American Mathematical Society, the American Association of Physics Teachers, the National Academy of Science, the American Academy of Arts and Sciences, the Royal

Netherlands Academy of Sciences and Letters, the American Association for the Advancement of Science, the Austrian Academy of Sciences, he is corresponding member of the Gesellschaft der Wissenschaften, Göttingen, and foreign member of the Royal Society of Great Brittain. He was a member of the General Advisory Committee to the U.S. Atomic Energy Commission from 1952–1957, was reappointed to this committee in 1959 and served on it until 1964.

MARIA GOEPPERT MAYER

The shell model

Nobel Lecture, December 12, 1963

1. Models

There are essentially two ways in which physicists at present seek to obtain a consistent picture of atomic nucleus. The first, the basic approach, is to study the elementary particles, their properties and mutual interaction. Thus one hopes to obtain a knowledge of the nuclear forces.

If the forces are known, one should in principle be able to calculate deductively the properties of individual complex nuclei. Only after this has been accomplished can one say that one completely understands nuclear structures.

Considerable progress in this direction has been made in the last few years. The work by Brueckner[1], Bethe[2] and others has developed ways of handling the many-body problem. But our knowledge of the nuclear forces is still far from complete.

The other approach is that of the experimentalist and consists in obtaining by direct experimentation as many data as possible for individual nuclei. One hopes in this way to find regularities and correlations which give a clue to the structure of the nucleus. There are many nuclear models, but I shall speak only of one and leave the others to the next lecture by Professor Jensen.

The shell model, although proposed by theoreticians, really corresponds to the experimentalist's approach. It was born from a thorough study of the experimental data, plotting them in different ways and looking for interconnections. This was done on both sides of the Atlantic ocean, and on both sides one found that the data show a remarkable pattern. This pattern emerges if one plots properties against either the number of neutrons, or the number of protons in the nuclei, rather than against the mass number.

2. Magic numbers

One of the main nuclear features which led to the development of the shell structure is the existence of what are usually called the magic numbers. That

such numbers exist was first remarked by Elsasser[3] in 1933. What makes a number magic is that a configuration of a magic number of neutrons, or of protons, is unusually stable whatever the associated number of the other nucleons. When Teller and I worked on a paper on the origin of elements, I stumbled over the magic numbers. We found that there were a few nuclei which had a greater isotopic as well as cosmic abundance than our theory or any other reasonable continuum theory could possible explain. Then I found that those nuclei had something in common: they either had 82 neutrons, whatever the associated proton number, or 50 neutrons. Eighty-two and fifty are «magic» numbers. That nuclei of this type are unusually abundant indicates that the excess stability must have played a part in the process of the creation of elements.

My attention was then called to Elsasser's papers written in 1933. In the year 1948 much more was known about properties of nuclei than was available to Elsasser. The magic numbers not only stood up in the new data, but they appeared more clearly than before, in all kinds of nuclear processes. It was no longer possible to consider them as due to purely accidental coincidences.

The magic numbers, as we know them now are:

2, 8, 20, 28, 50, 82, 126

and most importantly, they are the same for neutrons and protons. Fig. 1 shows the magic numbers and below them the stable nuclei containing magic number of protons or of neutrons.

Sn, $Z = 50$ is the element with the largest number of stable isotopes, namely 11. There are 6 stable nuclei with 50 neutrons, and 7 with 82 neutrons, whereas normally there are only 2 or 3 nuclei with the same number of neutrons.

It has long been known that helium, with two neutrons and two protons, is very tightly bound. An extra nucleon cannot be attached to the helium core, that is ^{5}Li and ^{5}He do not exist. The number 8 is encountered at $^{16}_{8}O_8$. It takes an unusual amount of energy to remove a neutron from this nucleus. On the other hand, the ninth, the extra neutron beyond the 8–8 shell, in $^{17}_{8}O_8$, is very weakly bound.

For nuclei heavier than ^{40}Ca the number of protons is less than that of neutrons and only then does it become clear that the stability is connected with the neutron number or the proton number, and not with the total number of both.

Let me show just two examples. The first one is taken from the work of Suess and Jensen[4], and is derived from the energy changes in β decay. Fig. 2

shows the energy difference between pairs of isobaric nuclides with neutron excess 3 and 1, with the common mass number as abscissa. The light nuclides, for which the energy difference is positive decay by β^- emission to the nuclides with $N-Z=1$. For the heavier nuclides, the neutron excess of 3 is the stable isobar, the energy is negative.

Magic number nuclides

Number of protons	2	8	20	28	50	82	126
	^{4}He	^{16}O	^{40}Ca	^{58}Ni	^{112}Sn	^{204}Pb	
		^{17}O	^{42}Ca	^{60}Ni	^{114}Sn	^{206}Pb	
		^{18}O	^{43}Ca	^{61}Ni	^{115}Sn	^{207}Pb	
			^{44}Ca	^{62}Ni	^{116}Sn	^{208}Pb	
			^{46}Ca	^{64}Ni	^{117}Sn		
			^{48}Ca		^{118}Sn		
					^{119}Sn		
					^{120}Sn		
					^{122}Sn		
					^{124}Sn		
Number of neutrons	2	8	20	28	50	82	126
	^{4}He	^{15}N	^{36}S	^{48}Ca	^{86}Kr	^{136}Xe	^{208}Pb
		^{16}O	^{37}Cl	^{50}Ti	^{87}Rb	^{138}Ba	^{209}Bi
			^{38}A	^{51}V	^{88}Sr	^{139}La	
			^{39}K	^{52}Cr	^{89}Y	^{140}Ce	
			^{40}Ca	^{54}Fe	^{90}Zr	^{141}Pr	
					^{92}Mo	^{142}Nd	
						^{144}Sm	

Fig. 1. The magic numbers.

One would expect to find a smooth curve, sloping downward. Except for one point it is indeed so. This point is ^{39}A with 21 neutrons and 18 protons. A smooth interpolation of the curve would predict ^{39}A is stable, and that its isobar ^{39}K is unstable against β^+ emission. However, ^{39}A is unstable against β^- emission by about 0.5 MeV. The explanation of this anomaly is the low binding energy of the 21st neutron in ^{39}A, while the 19th proton into which it is transformed has the higher binding energy of the proton shell which closes at 20. That the energies drop again sharply is due to the fact that now $Z=20$ is involved.

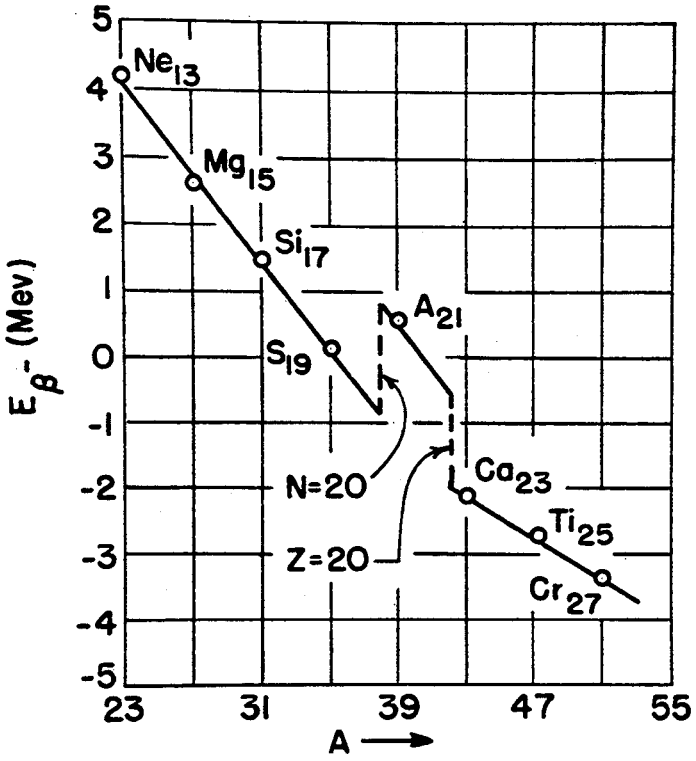


Fig. 2. Beta decay energies in the neighborhood of $N = 20$.

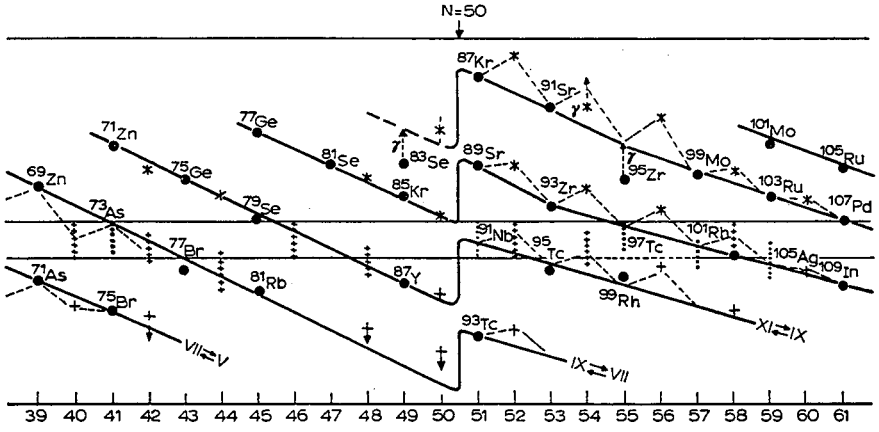


Fig. 3. Beta decay energies in the neighborhood of $N = 50$.

These types of discontinuity occur at all magic numbers. Fig. 3 shows it at the magic number $N = 50$, where it occurs for various numbers of the neutron excess.

The other example is that of the highest magic number 126, which occurs only for neutrons, and which was noticed long ago. Again, the prediction is that it would be easy to remove the 127th or 128th neutron, but that it takes a considerable amount of energy to remove the 126th or 125th neutron, whatever the associated proton number. Fortunately, this is the region in which α-decay occurs, in which two neutrons are lost by the nucleus, along with two protons. And the prediction is simply born out by the facts.

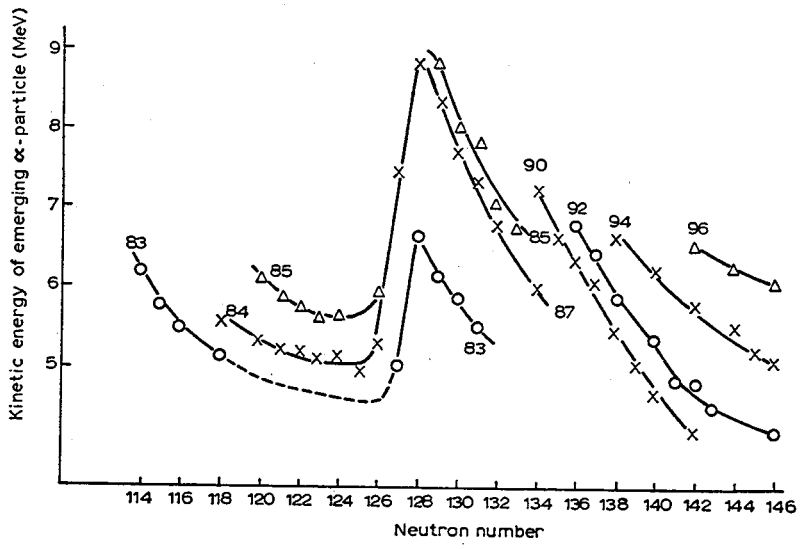


Fig. 4. Energy release in alpha decays.

Fig. 4 (after Seaborg and Perlman) shows the experimental data of the kinetic energy of the emerging α-particle, with the number of neutrons as abscissas. Isotopes of the same elements are connected by lines. The trend of the curves for the neutron-rich nuclei is easy to understand. But for all elements the energy reaches its peak at 128 neutrons, and then drops sharply when the 126th and the 125th neutron is removed from the nucleus.

From those and similar data one can estimate that the discontinuity of the binding energy at the magic numbers is about 1.5 to 2 MeV.

3. The atomic analogue

The strong binding of a magic number of nucleons and weak binding for one more, immediately brings to mind a similar, only relatively much stronger effect which occurs in the electronic structure of atoms. The energy required to remove an electron from an atom is measured by the ionization potential. The closed electron shells occur in the noble gas atoms, which have a very high ionization potential. The atoms with atomic number larger by one unit, the alkali, have a very low ionization potential. For instance, for argon with atomic number 18 and 18 electrons, the energy to remove one electron is 15.69 eV, whereas the energy to remove the 19th electron from potassium is only 4.32 eV. That is, the binding energy of the last electron in argon is about three and a half times that in potassium. In the nuclear cases, the change in binding energy across a magic number is only two MeV out of about an average value of six, which is only about thirty to forty per cent. Yet the experimental facts leading to magic numbers were sufficiently marked and they could hardly arise from accident. It seemed to be worthwhile to attempt to explain them in the same way as the noble gases. Indeed one might try to copy the essential features of the atomic structure for nuclear structure.

The simplest atom is hydrogen, in which one electron is subjected to the spherically symmetrical attraction of one proton. The quantum mechanical levels are characterized by two numbers, of which one, n, is called the principal quantum number. The other one, l, determines the angular momentum. By accident, due to the fact that the potential is proportional to the reciprocal of the distance, the energy depends only, or almost only, on the principal quantum number n.

Classically, in a field of spherical symmetry the angular momentum is a constant of the motion. In quantum mechanics, the orbital angular momentum is quantized, so that its magnitude in units of Planck's constant $\hbar$ is an integral value l. A level of given l contains $2l+1$ discrete states of different orientation in space, characterized by an integer m_l with $-l \leq m_l \leq l$. These numbers give the projection of the angular momentum on some axis in space. The states of given l and different values of m_l always have the same energy in any potential of spherical symmetry, even with potentials other than r^{-1}.

It is customary, to designate the levels of different l by letters in the following way:

$l=$	0	1	2	3	4	5
	s	p	d	f	g	h

Finally, the electrons have an intrinsic spin of 1/2 about their own axes which can only have two directions in space. The direction of the spin can be described by a quantum number m_s, with $m_s = +1/2$ for spin «up» and $m_s = -1/2$ for spin «down». Thus every one of the $2l+1$ states of given l is now double.

The basic assumption for the explanation of the periodic table is the following: Considering one particular electron, arbitrarily chosen, we shall assume that the action on it of all other electrons, as well as of the nucleus, can be approximated by a spherically potential $V(r)$. Since this potential is no longer proportional to the reciprocal distance the levels in it will be shifted, compared to hydrogen, and in such a way that the energy now depends on the angular momentum, measured by l, which is still quantized. The structure of the periodic system then follows from the Pauli principle: A quantum state of given n, l, m_l, m_s, can be occupied by only one electron. In other words, an energy level characterized by l can be occupied by no more than $2(2l+1)$ electrons. One builds up the periodic table by increasing the nuclear charge Ze and with this the number of electrons Z. To get the ground state of the atom we have to fill the lowest individual electron levels successively with as many electrons as the Pauli principle permits. When two successive levels are far apart in energy, we speak of closing an atomic shell at the element for which the lower of these is filled. At the next element the next electron can only be brought into the atom at a much higher level, with much less binding energy.

This description of atomic structure may be termed the individual orbit model.

4. Individual orbits in the nucleus

In analogy with atomic structure one may postulate that in the nucleus the nucleons move fairly independently in individual orbits in an average potential which we assume to have spherical symmetry. The value of the angular momentum, l, is quantized and contains $(2l+1)$ states, $-l \leq m \leq l$.

The assumption of the occurrence of clear individual orbits of neutrons and protons in the nucleus is open to grave doubts. In the atom, there is firstly the dominant attraction of the nucleus. The Coulomb repulsion between the electrons is of long range, so that the potential acting on one electron does not depend sensitively on the precise position of the others. In the nucleus, on the other hand, the forces are of short range, so that the potential on one nucleon should depend strongly on the position of the others. In other words, one

would expect that a nucleon collides with another one long before it has traversed its orbit even once.

Actually, perturbation by collisions is not as severe as one would at first expect, since the Pauli principle forbids collisions that deflect nucleons into already filled orbits, and therefore most of the intuitively expected collisions do not occur. We shall pursue the description of the nucleus by the independent orbit model. It still remains surprising that the model works so well.

There are several differences between the nucleus and the electrons in the atom. Firstly, the average potentials in the two cases are quite different. Thus the atomic shells numbers and the nuclear magic numbers will be entirely different from each other. One expects that the average nuclear potential has the form of a trough in three dimensions, where the potential is negative and rather constant inside the nucleus, rising abruptly to zero at the edge.

The second difference is that the nucleus contains two kinds of particles, neutrons and protons, each with intrinsic spin 1/2. We shall assume that the nuclear potential is the same for protons and neutrons. This assumption is now known to be in agreement with the evidence of many high-energy experiments, but at the time of the nuclear shell model development it was supported most strongly by the fact that the magic numbers were the same for neutrons and protons. The Pauli principle requires that just as in the case of electrons, a level of given l, can be occupied by no more than $2(2l+1)$ nucleons *of one kind.*

In a potential trough the lowest level is 1 s, $l=0$ with room for two neutrons and two protons. Two protons and two neutrons in this level make ^{4}He. The next level is 1 p, $l=1$, which has 6 states so that the 1 s and 1 p level together have room for 8 nucleons of one kind. Since there are two kinds, neutrons and protons, altogether 16 nucleons can be accomodated, leading to ^{16}O. Thus the uniquely stable numbers are easily explained for the light nuclei.

This is by no means new, but based on Wigner's pioneering work[5] on the light nuclei. Wigner's theory is able to explain with good approximation all the properties of light nuclei, spins, magnetic moments, transition probabilities, etc.

Its natural extension, however, failed in predicting the properties of heavy nuclei, and somehow, the theory of individual orbits in the nucleus went out of fashion. But nobody who has read Wigner's articles will ever forget them.

Fig. 5 shows some types of average potentials, a square well in 3 dimensions, a well with rounded edges, and a three-dimensional harmonic oscillator. The three-dimensional oscillator has equally spaced levels, which are highly degenerate, but which split up into several levels of different angular momentum

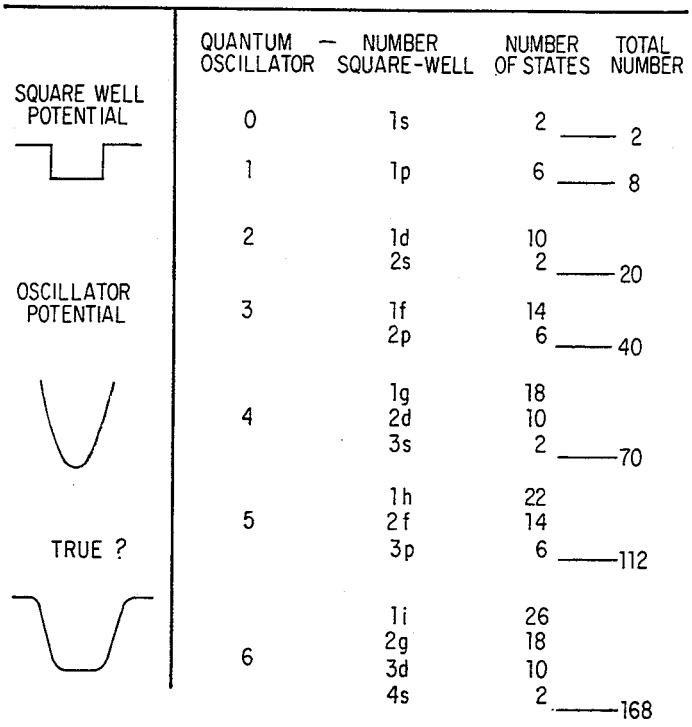


Fig. 5. Energy levels in a square-well.

l in the square well. I shall frequently use the term «oscillator shell», by which I mean the group of levels which for the harmonic oscillator would have the same energy. All levels of one oscillator shell have the same parity, that is they contain either only odd, or only even values of l.

The right-hand side shows the order of the levels with different values of l, and the number of nucleons of each kind which fill these levels, in agreement with the Pauli principle.

The magic number 8 corresponds to filling all levels up to the oscillator shell $n=1$. The magic number 20 is still explained as filling the oscillator shells up to $n=2$. But beyond that the system breaks down. There is no experimental trace of a gap in the level system at the oscillator shell numbers of 40, 70 and 112, and no reason seen for the observed gaps at 28, 50 and 82, and 126. Actually, for a potential which has not the oscillator shape, but is a «square well» in character the gap in energy at the oscillator shells is no longer marked. The answer was that we had copied the atomic analogue too closely.

5. *Nuclear shells*

Elsasser had tried to explain the magic numbers by assuming that the nuclear potential in heavier nuclei is quite different from a square well. Subsequent work showed quite conclusively that a change in the shape of the potential, even a change which was quite unrealistic could not explain the magic numbers. It was kind of a jigsaw puzzle. One had many of the pieces (not only the magic number), so that one saw a picture emerging. One felt that if one had just one more piece everything would fit. The piece was found, and everything cleared up.

At that time Enrico Fermi had become interested in the magic numbers. I had the great privilege of working with him, not only at the beginning, but also later. One day as Fermi was leaving my office he asked: «Is there any indication of spin–orbit coupling?» Only if one had lived with the data as long as I could one immediately answer: «Yes, of course and that will explain everything.» Fermi was skeptical, and left me with my numerology.

I do not know how many false starts my German colleagues made, but I had certainly made many. This one was not. The magic numbers from 28 on can definitely not be obtained by any reasonable extrapolation from the lower numbers, but form a different sequence. There are two different series of numbers, 2, 8, 20, 40..., of which 40 is no longer noticeable, and another, 6, 14, 28, 50, 82, 126 of which the first two at 6 and 14 are hardly noticeable. The second series is due to spin–orbit coupling. In ten minutes the magic numbers were explained, and after a week, when I had written up the other consequences carefully, Fermi was no longer skeptical. He even taught it in his class in nuclear physics.

At about the same time Haxel, Jensen and Suess had the same idea.

Let me explain what spin–orbit coupling, or more correctly, coupling of spin and orbital angular momentum means. Earlier I have spoken somewhat vaguely about the quantum number of the intrinsic spin, m_s, which is $+1/2$ for «spin up» and $-1/2$ for «spin down». Up and down with respect to what? If one has just one nucleon in a shell the only preferred direction is that of the orbital angular momentum. So spin, which is an angular momentum, can be parallel or anti-parallel to the orbital angular momentum. The total angular momentum has then the magnitude of $j=l+1/2$ or $j=l-1/2$. The number of states in each of two levels is $2j+1$ due to differing orientation of the total angular momentum. There is no longer a factor 2, since the spin is now fixed. Notice that $[2(l+1/2)+1]+2(l-1/2)+1=2(2l+1)$, so that there are still

the same total number of states. I shall refer to the half integer j of a nucleon in a given state as its spin in this state.

The basic assumption of the shell model is that there is a strong interaction between spin and orbital angular momentum, giving the level $j = l + 1/2$ a considerably lower energy. Since the splitting is proportional to l, and presumably goes down somewhat with nuclear size, prominant gaps in the level structure will always occur when a high orbital angular momentum occurs for the first time. This explains the magic numbers. Let me show how this works for the number 28. The oscillator shell closes at 20. The next levels are 1 f ($l = 3$) and 2 p ($l = 1$), in that order. The 1 f level splits into $j = 7/2$ and $j = 5/2$,

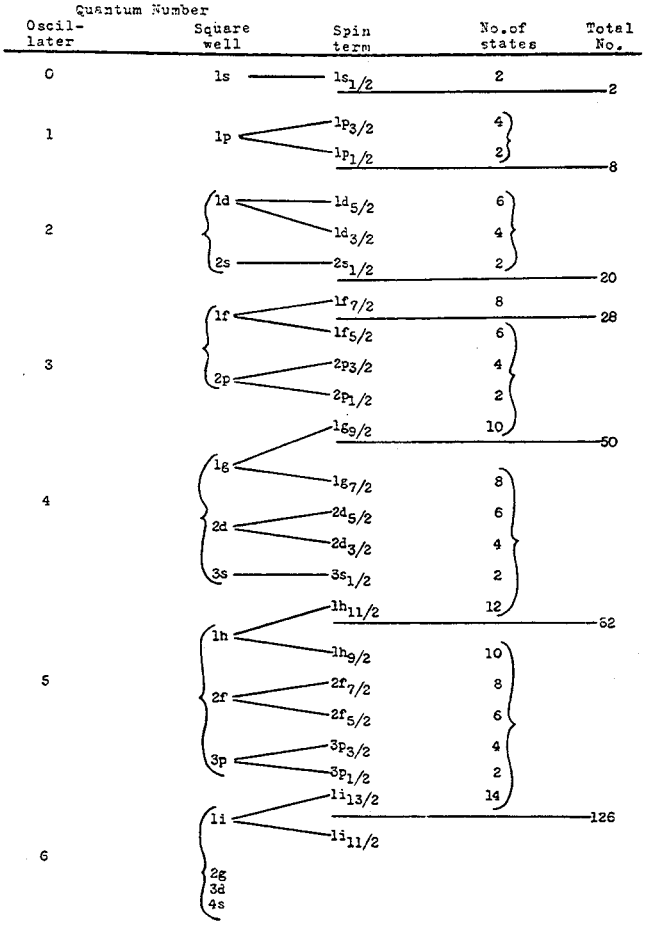


Fig. 6. Schematic level diagram.

with 7/2 lower. Since the energy difference is large, and the 7/2 level contains 8 states, we find the gap at 20+8=28 nucleons. All the magic numbers are explained in the same way. And since they are explained and no longer magic, I shall from here on call them shell numbers.

The assumption of a strong spin–orbit coupling contradicted the earlier tradition which assumed that spin–orbit coupling was very weak. Our attitude was «We know so little about nuclear forces.» By now, there is ample evidence for the fact that spin–orbit interaction in nuclei is indeed an important effect. Fig. 6 shows a very schematic level scheme. At the left side are the numbers and levels of the oscillator shell. In the right-hand side is the level

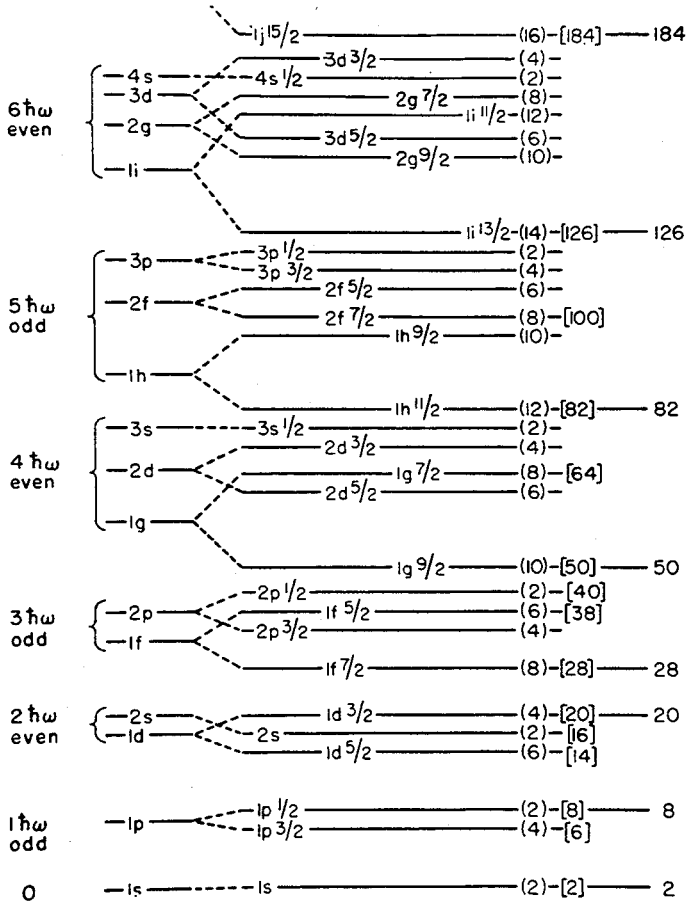


Fig. 7. Realistic level diagram for protons.

scheme with strong spin–orbit coupling. A magic number of neutrons or protons is obtained when the states of all oscillator shells up to a given one are filled with one each, and in addition the level of highest spin of the next oscillator is also filled with its complement of $2j+1$ nucleons.

Fig. 7 shows a fairly realistic level scheme for protons. It shows the fairly small splitting of the 1 p or $l=1$ level. The splitting of the 1 f ($l=3$) 1 g ($l=4$) 1 h ($l=5$) are increasingly larger. Within the shells the level order is harder to predict. It depends on the relative strength of spin–orbit coupling and the deviation from the oscillator potential. The detailed order in which we put levels is dictated by experiment. For instance, in the shell with oscillator number 3 we find that the 29th proton, after the 7/2 shell is filled, is in a 3/2 orbit. So the level p ($l=1$) 3/2 is lower than the level f ($l=3$) 5/2, the partner to the 7/2 state.

For neutrons, the level scheme is the same as for protons for the light nuclei up to neutron number 50. Above this, the Coulomb energy makes itself felt. It has the effect that the repulsion of protons favors orbits with higher angular momentum. Thus for neutrons, for instance, the 51st neutron is in the d level of $j=5/2$, whereas the proton is in the g level of $j=7/2$. This effect is never large enough to effect the shell number.

6. Predictions of the shell model

To be a reasonable model of nuclear structure the shell model must be able to explain and predict other nuclear properties than just a half dozen numbers. It is indeed able to do this.

Let me first consider the angular momenta, or nuclear spins, not of the individual nucleons but of the whole complex nuclei, which I shall designate by capital J. Hundreds of these have been measured. A closed shell, or a filled level, has angular momentum zero, since all states of different direction of the angular momentum contain one nucleon. Hence, nuclei with one nucleon outside (or one nucleon missing from) a closed shell of neutrons and of protons, or even of filled levels of both, should have a nuclear spin corresponding to the level of the single last nucleon and the spin of the individual particle orbit is predicted by the shell model. This is quite a severe test, since we find there would be no possible way to explain a disagreement with the model. Happily, all known nuclei of this type have indeed the predicted spin and parity.

Examples are ${}^{17}_{8}O_9$ with one nucleon outside the doubly closed shell of ${}^{16}_{8}O_8$, which has a spin of 5/2 and positive parity, which is the prediction. Another is ${}^{209}_{83}Bi_{126}$, which has one nucleon outside the closed shells of 82 protons and 126 neutrons and has a spin 9/2, in agreement with the predictions.

In nuclei where both neutrons and protons fill shells incompletely, the individual nucleons add their spin vectors to a total spin vector $\vec{J}$. Even with the restriction of the Pauli principle, very many states of total angular momenta exist. For instance, if there are three identical nucleons in the 7/2 shell there are 6 levels of different magnitude of total angular momentum, ranging from 3/2 to 15/2. It is very fortunate that of the vast number of complicated levels only the simplest ones occur as the ground state of nuclei.

There are further regularities. For instance, in bismuth there exist 5 isotopes of odd mass number in which the neutron number is even. All have a nuclear spin measured to be 9/2, namely, that of the 83rd proton. Thus the even numbers of neutrons, ranging in this case from 116 to 126 do not influence the spin.

Another example is the region where the first 7/2 shell is being filled. Here we know the spins of 8 nuclei with an even number of protons and odd number of neutrons ranging from 21 to 27. Seven of these have nuclear spins 7/2, one has 5/2. There are also 5 nuclei with an even number of neutrons and an odd number of protons ranging from 21 to 27, of which 4 have spins 7/2, one has 5/2. The numbers 21 to 27 correspond to 1, 3, 5, 7 nucleons in the 7/2 shell. So for nuclei in which both neutrons and protons fill shells incompletely, there emerge rules by which one may predict how the individual nucleons couple their spins to the total nuclear spin J. In a nucleus with an even number of neutrons and odd number of protons the neutrons couple their spins to zero and do not influence the nuclear angular momentum. The protons usually couple their spins to a total angular momentum J which is equal to the angular momentum j of the level being filled, and only rarely less by one unit. The same statement holds when the words neutron and proton are interchanged.

These rules are sometimes expressed in a different way and lead to what is called the single-particle model. It is an experimental fact that all nuclei with an even number of neutrons and of protons have angular momentum zero. Thus, in a nucleus of even neutron number N, odd proton number Z, there is an even-even nuclear core with N neutrons and $Z-1$ protons. The last proton occupies an orbit around the spinless core, and this orbit is prescribed by the shell model. All nuclear properties, spin, magnetic moment, etc. are en-

tirely due to the last odd particle. Actually the shell model has never been proposed in quite so simple and naive a fashion.

These coupling rules, considerably less complex than those for atoms, do have some theoretical basis, namely a simplified calculation of energies predicts them. If one considers just several particles of the same kind in the same level j, and assumes that they interact with each other with a very short range force, one finds indeed that for an even number of nucleons the ground state has spin zero. For an odd number, the ground state has the spin J which is equal to the j of the level being filled. The eigen functions of $J=0$ for even, and of $J=j$ for odd particle number are those of lowest seniority.

With these rules we should be able to explain or predict the spins of all nuclei. Up to neutron or proton number a little above 50 this simple theory and experiment are in excellent agreement. Beyond this, there are very many levels in the shell 50–82 and these levels lie close together in energy, so that one can explain just about anything. Besides, nuclei with more than 90 neutrons are highly deformed, and the assumption of a potential with spherical symmetry is no longer the best starting point. This will be discussed in the next lecture. However as the closed shells $Z=82$ and $N=126$ are approached, there is no longer a large deformation, and the predicted and measured spins again agree.

Another quantum number which the model predicts is the parity. We not only predict the spin, but also the angular momentum l of each level. A level with odd l has odd parity, one with even l has even parity. Parity can be measured in various ways, and there is again complete agreement with the predictions.

Besides the ground states of nuclei one can also investigate the excited states. One type of excited states are the isomeric levels, which are levels of a very long lifetime, hours, days or even years. The explanation of this phenomenon is that the spins of the isomeric state and the ground state are very different, so that the return to the ground state by the emission of a light quantum is greatly hindered since the light quantum has to take up the difference in angular momentum. The transitions are not dipole but octupole or 2^4 pole transitions, which are very slow. In nuclei of odd mass number an excited state can be produced by raising the last odd nucleon into an adjacent higher level. Now there are only very definite regions where low and high spins are close in energy, namely at the end of the shells where the lowest angular momenta of one oscillator shell occur, and immediately above them the states of highest angular momentum of the next oscillator level. Thus, isomerism should occur

only if the number of last odd particle is between 38 and 50, or between 64 and 82, or between 100 and 126. In addition, the shell model predicts that all these transitions involve a change in parity. This is a rather strong statement and ties isomerism to the neutron or proton number.

Some of the best work on isomerism has been done in Sweden[4] and has led to one of the nicest confirmations of the shell model. The three regions of isomerism are now called islands of isomerism. Long-lived and low-lying isomeric levels in nuclei of odd A occur only in the three islands. If one considers the mass number only, no regularities appear, since different islands of proton isomerism and neutron isomerism overlap in mass numbers.

For instance $^{115}_{49}$In in the first island has an isomeric state with a half-life of 5.1 hours. This is due to the transition of a proton from the $j=1/2$ level to the ground state which has spin 9/2. For mass numbers higher by two, one finds $^{117}_{50}Sn_{67}$ with an isomeric state of half-life 14 days. This is due to the odd neutron, which goes from the excited level $j=11/2$ to a level $j=3/2$ which is expected to happen in the second island.

7. Failures of the shell model

After all this praise of the shell model, it is high time to emphasize its shortcomings. Even a crude nuclear model should be able to explain quantum numbers, like the spin, which is either integer or half integer, but never in between, or parity, which is either even or odd. The shell model, as I have presented it, can indeed do this, and in this form has the advantage that it can explain or predict these quantum numbers for most nuclei.

However, the single-particle model, namely the rule for the coupling of the spins of individual nuclei, which essentially postulate that everything depends only on the last odd nucleon can be at best a very rough approximation to the truth. This becomes obvious when one tries to calculate nuclear properties which are not integers but can be measured to seven significant figures. One would hope to get approximate agreement, say to 10%. Unfortunately, this is not so. For example, take the magnetic moments of nuclei. For a nucleus with odd proton, even neutron number, the magnetic moments, according to the shell model, should depend only on the state of the last odd proton and are easy to compute. For any value of the spin, we calculate two different values of the magnetic moment, for the two different values of l, $l=j-1/2$ and $l=j+1/2$. In Fig. 8 the magnetic moments of odd proton, even neutron

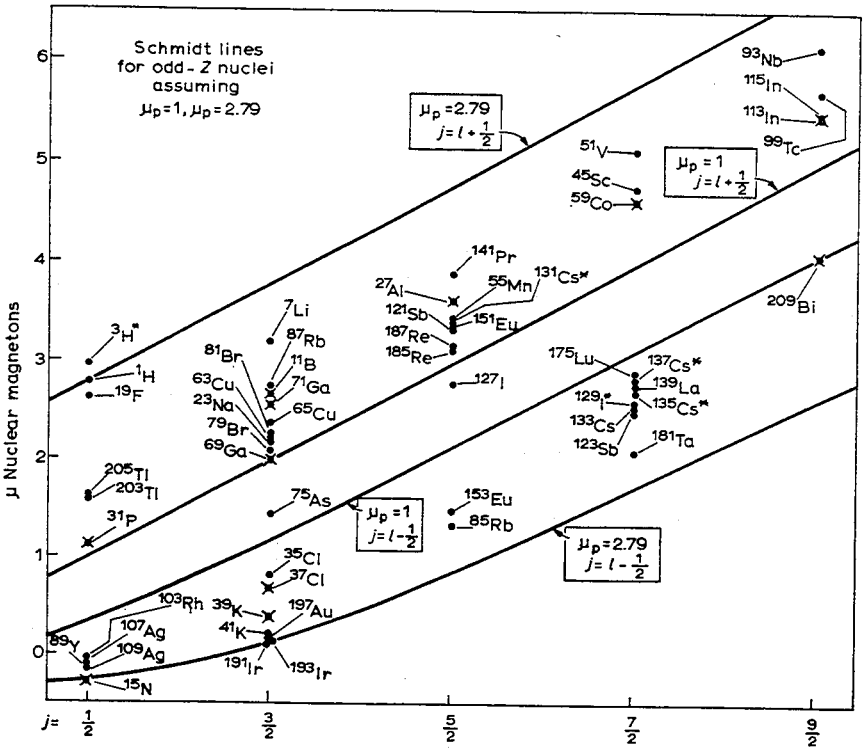


Fig. 8. Magnetic moments of nuclei with odd proton, even neutron number.

nuclei are plotted with the nuclear angular momentum as abscissa. The lines at the two extremes are the calculated ones. The middle lines are what one would obtain if the proton were a simple Dirac particle, and are added merely to emphasize the division into two groups. The difference between calculated and measured values are distressingly large. Only one general trend remains. The nuclei in the upper group, nearer to the line for $j = l + 1/2$ are indeed those for which we found that spin and orbital angular momentum are parallel, those in the lower group were assigned anti-parallel orientation.

This shows that much more careful calculations of the interaction between the nucleons are required to get better numerical agreement. For individual nuclei, or special groups of nuclei, such calculations have been made by many people using the shell model as first approximation, and different procedures to compute higher approximations. In particular, Talmi[7] has made great progress in developing a more refined shell model.

Finally, even the assumption of strong spin–orbit coupling is open to criti-

cism, at least for the light nuclei. For these the model can easily be refined by taking into account both protons and neutrons in the nucleus, and constructing eigen functions of lowest isotopic spin. One should compare the results obtained to those of Wigner's calculations[5]. Although Wigner also used the independent particle model, his method is in some sense the direct antithesis to the shell model. In Wigner's theory, spin–orbit coupling is assumed to be very weak, whereas in the shell model spin and orbital angular momentum are assumed to be rigidly coupled.

Actually, Wigner's values for all nuclear properties agree better with the experimental results for the light nuclei. It seems that the truth is in the middle, spin–orbit coupling is present, but not predominant. The calculation for «intermediate» coupling are more involved than either extreme, but they have been done by many people[8] for different nuclei, and have led to much closer agreement between theory and experiment.

The shell model has initiated a large field of research. It has served as the starting point for more refined calculations. There are enough nuclei to investigate so that the shell modellists will not soon be unemployed.

1. K. A. Brueckner, A. M. Lockett and M. Rotenberg, *Phys. Rev.*, 121 (1961) 255.
2. H. A. Bethe and J. Goldstone, *Proc. Roy. Soc.(London)*, *Ser. A*, 238 (1957) 551.
3. W. Elsasser, *J. Phys. Radium*, 4 (1933) 549.
4. H. E. Suess and J. H. D. Jensen, *Arkiv Fysik*, 3 (1951) 577.
5. E. Feenberg and E. Wigner, *Phys. Rev.*, 51 (1937) 144.
6. E. Wigner, *Phys. Rev.*, 51 (1937) 947. See, for instance, references in M. Goldhaber and A. W. Sunyar, *Phys. Rev.*, 83 (1951) 906.
7. A. de-Shalit and I. Talmi, Academic Press, New York, 1963.
8. For reverences see article by D. Kurath, in K. Siegbahn (Ed.), *Alpha, Beta and Gamma Ray Spectroscopy*, Vol. 1, North Holland, Amsterdam, 1965, p. 583.

Biography

Maria Goeppert Mayer was born on June 28, 1906, in Kattowitz, Upper Silesia, then Germany, the only child of Friedrich Goeppert and his wife Maria, née Wolff. On her father's side, she is the seventh straight generation of university professors.

In 1910 her father went as Professor of Pediatrics to Göttingen where she spent most of her life until marriage. She went to private and public schools in Göttingen and had the great fortune to have very good teachers. It somehow was never discussed, but taken for granted by her parents as well as by herself that she would go to the University. Yet, at that time it was not trivially easy for a woman to do so. In Göttingen there was only a privately endowed school which prepared girls for the «abitur», the entrance examination for the university. This school closed its doors during the inflation, but the teachers continued to give instructions to the pupils. Maria Goeppert finally took the abitur examination in Hannover, in 1924, being examined by teachers she had never seen in her life.

In the spring of 1924 she enrolled at the University at Göttingen, with the intention of becoming a mathematician. But soon she found herself more attracted to physics. This was the time when quantum mechanics was young and exciting.

Except for one term which she spent in Cambridge, England, where her greatest profit was to learn English, her entire university career took place in Göttingen. She is deeply indebted to Max Born, for his kind guidance of her scientific education. She took her doctorate in 1930 in theoretical physics. There were three Nobel Prize winners on the doctoral committee, Born, Franck and Windaus.

Shortly before she had met Joseph Edward Mayer, an American Rockefeller fellow working with James Franck. In 1930 she went with him to the Johns Hopkins University in Baltimore. This was the time of the depression, and no university would think of employing the wife of a professor. But she kept working, just for the fun of doing physics.

Karl F. Herzfeld took an interest in her work, and under his influence and

that of her husband, she slowly developed into a chemical physicist. She wrote various papers with Herzfeld and with her husband, and she started to work on the color of organic molecules.

In 1939 they went to Columbia. Dr. Goeppert Mayer taught one year at Sarah Lawrence College, but she worked mainly at the S. A. M. Laboratory, on the separation of isotopes of uranium, with Harold Urey as director. Urey usually assigned her not to the main line of research of the laboratory, but to side issues, for instance, to the investigation of the possibility of separating isotopes by photochemical reactions. This was nice, clean physics although it did not help in the separation of isotopes.

In 1946 they went to Chicago. This was the first place where she was not considered a nuisance, but greeted with open arms. She was suddenly a Professor in the Physics Department and in the Institute for Nuclear Studies. She was also employed by the Argonne National Laboratory with very little knowledge of Nuclear Physics! It took her some time to find her way in this, for her, new field. But in the atmosphere of Chicago, it was rather easy to learn nuclear physics. She owes a great deal to very many discussions with Edward Teller, and in particular with Enrico Fermi, who was always patient and helpful.

In 1948 she started to work on the magic numbers, but it took her another year to find their explanation, and several years to work out most of the consequences. The fact that Haxel, Jensen and Suess, whom she had never met, gave the same explanation at the same time helped to convince her that it was right. She met Jensen in 1950. A few years later the competitors from both sides of the Atlantic decided to write a book together.

In 1960 they came to La Jolla where Maria Goeppert Mayer is a professor of physics. She is a member of the National Academy of Sciences and a corresponding member of the Akademie der Wissenschaften in Heidelberg. She has received honorary degrees of Doctor of Science from Russel Sage College, Mount Holyoke College and Smith College.

They have two children, both born in Baltimore, Maria Ann Wentzel, now in Ann Arbor, and a son, Peter Conrad, a graduate student of economics in Berkeley.

J. HANS D. JENSEN

Glimpses at the history of the nuclear structure theory

Nobel Lecture, December 12, 1963

During the last weeks, I have often thought of my teachers, especially of the one man who had great influence on my attempts to gain some understanding of nuclei, Niels Bohr. I think it is also appropriate at this occasion to consider first the background from which our concepts of nuclear structure emerged.

I can devote only a few sentences to the time preceding Chadwick's discovery of the neutron (1932). At that time our information regarding the nucleus was very sparse. All we had was a chart of known stable isotopes with nuclear masses which were not very accurate, a few nuclear spins, an estimate of the nuclear radius, about $1.4 \cdot 10^{-3} A^{1/3}$, the phenomena of natural radioactivity, and a few known nuclear reactions. Ideas on nuclear structure were still dominated by Prout's hypothesis of 1815, that electrons and protons, the only known elementary particles, are bound together in a nucleus in such a way that A protons and $A-Z$ electrons form a nucleus of charge Z. But from the point of view of quantum mechanics a great puzzle was inseparably inherent in this picture. Consider the deuteron as the simplest example. According to the model, the deuteron contains two protons and one electron, just like the ion of the hydrogen molecule. Yet in the deuteron the linear dimensions are 10^{-5} times smaller than in the hydrogen molecule. The uncertainty principle requires very strong forces to confine electrons to such a small volume. These non-Coulomb forces should then show up just as well in the hydrogen spectrum and change the Balmer formula; in particular, they should give rise to a much larger splitting than that discovered later by Lamb. I cannot discuss other similarily grave inconsistencies of the model in this limited time.

In view of these conflicts many physicists, including Niels Bohr, were inclined to expect far-reaching changes in our basic physical concepts, even in quantum mechanics*. At that time one was tempted to consider alpha particles

* Some physicists thought that it might even become necessary to give up the conservation laws in their current form, especially in connection with the problem of beta decay.

as basic building blocks of nuclei. However, from those days a warning from Schroedinger still persists in my mind. During the late twenties he chided the participants in a Berlin seminar for their lack of imagination. In his impulsive manner he said: «Just because you see alpha particles coming out of the nucleus, you should not necessarily conclude that inside they exist as such.» And he gave an illustrative example from every-day life to show how such reasoning can lead to fallacious conclusions.

It is remarkable that very little information about nuclear structure could be gained from the study of alpha decay. Max von Laue has pointed this out very clearly in a letter to Gamow in 1926; he congratulated Gamow on his explanation of the Geiger–Nuttal law* in terms of the tunnelling effect and then went on: «however, if the alpha decay is dominated by quantum phenomena in the region outside the nucleus, we obviously cannot learn much about nuclear structure from it.». Gamow says that at first he was quite perplexed while reading these lines, but thinking it over he had to agree with von Laue. The situation that very little insight into nuclear structure could be gained from this oldest nuclear phenomenon persisted for a long time. Only about 6 years ago some progress was made when Mang applied the shell model to the problem of alpha decay. It seems to me that Mang's results justify Schroedinger's scepticism; the alpha particles obviously only form while emerging from the nucleus.

The discovery of free neutrons changed the situation entirely. Now it became possible to separate the grave difficulties of «the localization of electrons in the nucleus», to which I shall return later, from the specific problem of nuclear structure. Thus, in spite of Schroedinger's warning (this time, of course, regarding the neutrons), one could consider the hypothesis that protons and neutrons are the fundamental units within the nucleus. (Rutherford had already suggested this in conversations before Chadwick's discovery, and Harkins had published the same proposal). Specific nucleon–nucleon forces acting between them must be responsible for the nuclear binding. Heisenberg was the first to explore the consequences of this hypothesis, and to arrive at important concepts and results in a series of pioneering papers in the *Zeitschrift für Physik* (1932–1933).

These ideas can be separated into two stages. First, the saturation phenome-

* That is, the fact that the lifetime of an α-emitter changes by 25 powers of ten when the alpha-particle energy increases only by a factor of two.

non can be accepted as an empirical fact, *i.e.* the approximate proportionality of nuclear binding energy to the particle number A, as well as the proportionality of the nuclear volume to A, with a radius already mentioned. The numerical value of r was a crude estimate at that time; now we know from the Stanford experiments that it is about 20% smaller. These facts as well as the results of scattering experiments led to the conclusion that nuclear forces must have a short range. In spite of this shortness of range, in one of his papers Heisenberg considered the nucleus as a superposition of two Fermi gases (a neutron gas and a proton gas) which freely permeate each other and which by an averaged potential are confined to the given volume. The basic fact that stable nuclei have about the same number of neutrons and protons, $Z \approx A/2$, is explained on this basis as a consequence of Pauli's principle. In addition, one obtains the right order of magnitude for the curvature of the parabola defined by taking an $A =$ const. cross-section through the surface of binding energies of stable nuclei; the opening of the parabola was somewhat too large, with the new nuclear radius obtained by Hofstadter the agreement is even better. The decrease of the ratio $Z:A$ with increasing mass number is a natural consequence of the interplay between the accumulating Coulomb interaction and the consequences of Pauli's principle.

Thus the basic idea of the shell model was expressed for the first time, *i.e.*, the idea of free motion of individual nucleons in an averaged potential. Every further development was an inevitable extension of these ideas to a system with a finite number of particles*. The Leipzig school as well as Wigner and his co-workers devoted great effort to the study of light nuclei, mainly on the basis of the shell model. The particular stability of the nuclei ${}^{4}_{2}He_{2}$, ${}^{16}_{8}O_{8}$, and ${}^{40}_{20}Ca_{20}$, was not the only fact explained in this way. For example, Wigner and his co-workers came to a quantitative conclusion that the then unknown nuclides ${}^{36}_{16}S_{20}$ and ${}^{48}_{20}Ca_{28}$, should be stable; later these nuclides were in fact observed in mass spectrometers as natural isotopes with very small abundance.

* However, Heisenberg's interest extended far beyond this stage to the following question: By which properties of the forces can the nuclear saturation be explained? To account for this phenomenon, he introduced the concept of «exchange forces» which he formulated in terms of «isospin» formalism, first invented for this purpose. Thus he created a conceptual apparatus which is still used in discussing the most direct studies of nucleon–nucleon interaction, the scattering experiments. The quantitative results concerning exchange mixtures which would guarantee saturation are by now outdated. It is unfortunate that at that time one did not systematically pursue one other possible explanation of saturation: a property of the forces which is today usually called «hard core» or «almost hard core». Heisenberg also discussed this possibility in one of his papers.

Although this was somewhat a matter of luck in view of the insufficient knowledge of the forces, it was nevertheless one of the first predictions of nuclear theory to be verified experimentally. In 1937 Hund and Wigner, independently of each other, developed the concept of supermultiplets that played such an important role in classifying nuclides and in the systematics of beta decay. This concept was based on the assumption that nucleon–nucleon forces were essentially charge- and spin-independent. In the article by Bethe and Bacher in *Reviews of Modern Physics* (1936), which was soon called «Bethe's bible», very convincing arguments had been presented to show that, in fact, nuclear forces should not show much spin- and isospin-dependence; in particular the spin–orbit coupling should be very weak.

In the years immediately following the discovery of neutrons, a vigorous development of experimental nuclear physics began. This was partially due to the possibility of performing experiments with neutrons; partially to the completion of the first accelerators and to great improvements in measuring and counting techniques. For me these were the years of my first visits to Copenhagen and meeting Niels Bohr; in Copenhagen I was privileged to witness many attempts at a theoretical interpretation of the rapidly accumulating experimental data.

Two new phenomena were particularly important to the development of our concepts of nuclear structure: relatively high effective cross-sections for nucleon–nucleon scattering, and sharp, closely spaced resonances discovered by Fermi, Amaldi, and co-workers in slow-neutron scattering and capture. The latter phenomenon could not be explained at all in terms of the picture in which the neutron moves in an averaged potential. Thus Niels Bohr's concept of the «compound nucleus» originated. In his model, the state of the nucleus is characterized by an intimate coupling of all nucleons with each other; this description does not permit us to speak of the motion of a single nucleon independently of the simultaneous motion of all the others. However, this intuitive, semiclassical picture of Niels Bohr had to be brought into agreement with the postulates of quantum mechanics. To this day the golden bridge has been the Breit–Wigner formula. It originated independently outside Copenhagen, but it could soon be seen on every blackboard of Niels Bohr's institute. Naturally it received appropriate space in the above-mentioned «Bethe bible». Probably every theoretician has pondered long and often about its interpretation and about its proof; and it still occupies many minds.

One was inclined to describe even the ground state of a nucleus in terms of

Bohr's picture. A concept of nuclear matter, packed to saturation density and having a binding energy proportional to its volume became generally accepted. For finite nuclei a surface tension resulting from a surface energy proportional to the surface was to be added. The «Bethe bible» also contains an excellent discussion of the basis for these assumptions. The greatest success of this model was Bohr and Wheeler's theory of nuclear fission (1939), which contains almost everything that is understood to date (1963) about this phenomenon.

Schroedinger's remark, that one should not necessarily assume that the particles, observed as free particles emerging from the nucleus during nuclear transformations, must exist in the same form inside the nucleus, was emphasized in Fermi's paper on beta decay (1933–1934). In these papers the abovementioned dilemma, which arises from the concept of «electrons inside the nucleus or inside the neutron», was literally dissolved into nothing. Fermi drew radical consequences from the idea that the proton and the neutron are two quantum states of one single fundamental particle, the nucleon. Between these two states quantum transitions can take place. Such a transition is accompanied by the *creation* of an electron and a neutrino; (Fermi used Heisenberg's concept of isospin and its formalism in the theory of beta transformation). Today's young physicist, who already as a student juggles creation and annihilation operators on the blackboard, can hardly fathom the importance of the conceptual breakthrough contained in Fermi's theory. As an illustration, let me quote from a historical letter sent by Pauli to several friends and colleagues (December 1930) in which he proposed his neutrino hypothesis for the first time.

«... bin ich auf einen verzweifelten Ausweg verfallen,... nämlich die Möglichkeit, es könnten elektrisch neutrale Teilchen, die ich Neutrinos* nennen will, in den Kernen existieren... Das kontinuierliche β-Spektrum wäre dann verständlich unter der Annahme, dass beim β-Zerfall mit dem Elektron jeweils noch ein Neutrino emittiert wird...»

«I came to a desperate conclusion... that inside the nucleus there may exist electrically neutral particles which I shall call neutrinos*. The continuous beta spectrum becomes understandable if one assumes that, during beta decay, the emission of an electron is accompanied by the emission of a neutrino...»

* In his letter, written long before Chadwick's discovery, the word «neutron» appears instead of «neutrino»; the latter was adopted by Pauli later, following a suggestion by Fermi.

I emphasize the words *exist inside the nucleus* and *emission*. Pauli certainly did not choose these words simply to make his ideas more palatable to his experimental colleagues, but because the words characterize the physical concept of those days. It is all the more remarkable in view of the fact that the concept and mathematical technique of particle creation used by Fermi had been available long before in the so-called second quantization of Jordan, Klein, and Wigner. However, even in 1932, in his Handbuch article Pauli himself regarded that rather as a mathematical trick, it was Fermi's work which finally convinced him that there was tangible physics in it.

Yukawa's theory also became known in the middle of that decade. He emphasized that the forces between nucleons are transmitted by a field, which must show retardation effects, and quanta associated with the retardation effects, the mesons. The latter are perhaps of secondary importance in nuclear structure problems, since it was established by Heisenberg's investigations that in the nucleus the nucleons move so slowly that one may hope to understand the essential features of nuclear structure by using non-relativistic quantum mechanics. However, the strong coupling of the Yukawa field to its source is extremely important; its strength, $g^2/\hbar c$ is of the order of magnitude of ten (in contrast with the Sommerfeld constant $e^2/\hbar c = 1/137$ in electrodynamics). This led Niels Bohr to an idea on nuclear matter, which, to my knowledge, he never wrote down; but from conversations it has remained ineradicably engraved in my memory: since the field is strongly coupled to its sources, the hitherto existing picture of the «compound nucleus» may still be much too naive. Perhaps, the only sensible concept is to consider the whole nucleus as an «Urfeld» which is highly nonlinear because of such strong couplings. When this field is quantized, it must give (in addition to other conserved quantities, like angular momentum) integral charges Z, and energies (*i.e.* masses) that form a spectrum with values close to the integral numbers A, on which the «exaction energy» bands are superimposed. The assumption that inside the nucleus there exist Z protons and $(A-Z)$ neutrons such as we encounter them as free particles in appropriate experiments would then hardly make any sense.

Schroedinger's scepticism (mentioned at the beginning) would thus be formulated in its extreme. Nevertheless, Niels Bohr had thus hinted at a picture of the nucleus which closely resembles current concepts in high-energy physics on elementary particles and «resonances» (*e.g.*, such as the hyperons or the ϱ-, η-, etc. mesons). Certainly, one should not entirely forget such a point of view in nuclear physics either, although it has meanwhile been

shown*, that it is legitimate to speak of the existence of individual nucleons inside the nucleus in a useful approximation.

The picture of the nucleus just described is in accord with the fact that, just glancing at the table of stable isotopes, we can see that the nuclear properties are continuous functions of A and Z. However, there were indications of discontinuities and windings in the valley of the energy surface. I have already pointed out the exceptional cases of nuclei with Z and $N = 2, 8, 20$. It also seemed strange that the alpha energy does not increase uniformly as one goes further away from the alpha-stable nuclei in the mass valley; instead it is largest right at the polonium isotopes. This indicates that a special discontinuity occurs for $Z = 82$. In the diagram in which alpha energies are plotted against Z and N, we also see curves with steep slopes from $N = 128$ to $N = 126$; Gamow called this feature the «Heisen-Berg». The work of Seaborg and collaborators made the profile of these peaks even more striking. Elsasser, Guggenheim, Ivanenko, and others had attempted to explain these and other phenomena in terms of the shell model; however, it seemed impossible to accommodate the groups of numbers Z and $N = 2, 8, 20$, on the one hand and $Z = 82$, $N = 126$ on the other, under the same roof. Yet, mainly because of the success of Bohr's compound-nucleus model, there was a general tendency to consider these phenomena as curiosities of little significance to the fundamental questions of nuclear structure.

The war years and also the first few years thereafter brought the physicists in Germany into a stifling isolation, but at the same time they gave us some leisure to pursue questions off the beaten trails. At that time I had many discussions with Haxel in Berlin, later Göttingen, and with Suess in Hamburg on the empirical facts which single out the above-mentioned numbers. To Suess they became more and more significant, primarily in his cosmo-chemical studies: he found that in the interval between the numbers already mentioned, the numbers Z and $N = 50$ and $N = 82$ were also clearly prominent**. Haxel, at first quite independently, encountered the same numbers in the study of other nuclear data.

Although my two colleagues tried hard to convince me that these numbers might be a key to the understanding of nuclear structure, at first I did not know what to make of it. I thought the name «magic number», whose origin was

* In particular, through the work of Brueckner and recent literature inspired by it.

** V. M. Goldschmidt also came to the same conclusion; Suess and I had the privilege of discussing it with him in Oslo in 1942 and 1943.

unknown to me*, to be very appropriate. Then, a few years after the war, I had the privilege of returning to Copenhagen for the first time. There in a recent issue of the *Physical Review*, I found a paper by Maria Goeppert-Mayer, «On closed shells in nuclei», where she too had collected the empirical evidence pointing out the significance of the magic numbers. That gave me courage to talk about her work, along with our results, in a theoretical seminar. I shall never forget that afternoon. Niels Bohr listened very attentively and threw in questions which became more and more lively. Once he remarked: «But that is not in Mrs. Mayer's paper!»; evidently Bohr had already carefully read and pondered about her work. The seminar turned into a long and lively discussion. I was very much impressed by the intensity with which Niels Bohr received, weighed, and compared these empirical facts, facts that did not at all fit into his own picture of nuclear structure. From that hour on I began to consider seriously the possibility of a «demagification» of the «magic numbers».

At first I tried to remain as much as possible within the old framework. To begin with, I considered only the spin of the whole nucleus, since there appeared to exist a simple correlation between the magic nucleon numbers and the sequence of nuclear spins and their multiplicities. I first thought of the single-particle model with strong spin–orbit coupling** during an exciting discussion with Haxel and Suess, in which we tried to include all available empirical facts in this scheme. As we did this it turned out that, because of the spin–orbit coupling, the proton- and the neutron-number 28 should also be something like a magic number. I remember our being elated when we found some hints in the still meagre data that was available at that time. Nevertheless, I did not feel very happy about the whole picture, and I was not really surprised when a serious journal refused to publish our first letter, stating «it is not really physics but rather playing with numbers». Only when I thought of the lively interest in the magic numbers which Niels Bohr had shown did I dare send the same letter to Weisskopf who forwarded it to the *Physical Review*. Yet it was not until later, after I had presented our ideas in a Copenhagen seminar and been able to discuss them with Niels Bohr, that I finally gained some confidence. One of Bohrs first comments seemed remarkable to me: «Now I understand why nuclei do not show rotational bands in their spectra». With the accuracy of measurement available at the time, one had looked for such

* I learned only yesterday that the name was coined by Wigner.

** Fortunately, I was not too well versed in «Bethe's bible» and I did not remember the old arguments against a strong spin–orbit coupling too well.

spectra in lighter nuclei, which according to the liquid drop or a similar model should have relatively small moments of inertia and therefore widely separated rotational levels. As we know today, these light nuclei, like many others, in fact show no rotational bands; Bohr's argument was, of course, that in a picture in which single particles move independently in an average spherically symmetric potential, there was no place for a superimposed rotation of the nucleus as a whole, just as in the system of electrons in the atomic shells.

Even though the shell model finally proved to be more than just a convenient language in which the experimentalists could compare their notes, and although during the following years it led us to some understanding of a few fundamental features of nuclear structure, I still had to agree with Robert Oppenheimer when he told me: «Maria and you are trying to explain magic by miracles.» In a lecture at Oak Ridge, Wigner recently said something quite similar, of course more cryptic, in his careful way of choosing his words.

From the beginning it was clear to me as well as to Mrs. Goeppert-Mayer that the shell model could at best approximately describe the ground state and the low-lying excited states of nuclei. While the consequences of the Pauli principle for these states could possibly guarantee the self-consistency of our model, the Pauli principles becomes less and less stringent as the excitation energies become higher, and the nucleon–nucleon correlations arising from nuclear forces become increasingly important. In an exact description such correlations are, of course, also present in the ground state.

Therefore, during my next visit to Copenhagen I had a certain satisfaction when, questioned about news on the shell model, I could instead talk about the ideas which then occupied my namesake Peter Jensen and myself as well as Steinwedel and Danos. Following a suggestion by Goldhaber and Teller, we tried to provide a semi-classical explanation for the recently discovered large dipole absorption in the nuclear photoeffect at 15 to 20 MeV; that is, we described it as an excited state of nuclear matter in which all nucleons are in a state of motion with strict phase relations existing between all of them. In this way the frequency of the absorption maximum, as well as its dependence on the nuclear mass number, could be related in a satisfactory way to the symmetry energy and to the nuclear radius. The width of the «giant resonance» provided a measure of the rate at which such phase correlations disappear. Niels Bohr understood immediately why the study of this particular type of «collective motion» (in the present-day jargon of specialists) was of such great interest to me. Even though the importance of phase correlations be kept down in the

ground state by Pauli's principle, we wanted to determine at which excitation energies the correlations enforced by the nucleon–nucleon interaction become dominant over the effect of the averaged forces.

In the following years much work was devoted to the study of such correlations. A most remarkable feature of current nuclear physics was brought to light by the work of Kurath, of the former Harwell group (Flowers, Elliot, and others), and of the young Copenhagen School (Aage Ben Bohr, Mottelson, Nilsson and others). I mean the fact that, even though the groups started from points of view which, whilst complementing also limit each other, their quantitative results seem quite soon to meet, and to overlap, in the next steps of approximation.

The first group started from the shell-model point of view with a spherically symmetric potential, and handled the problem of correlations by calculating the configuration mixing which is caused by the forces acting individually in each pair of nucleons. Thus it was shown that, even with only a few nucleons outside a closed shell, one obtains level sequences very similar to rotational spectra. In this way, although it is difficult to perform a quantitative calculation, one can understand how in the case of nuclei with many nucleons outside closed shells (for example, the rare-earth region and the nuclei beyond) many close-lying and very different particle states can contribute to configuration mixing, creating such correlations that the ground state becomes a strongly deformed nucleus. The Copenhagen group started by treating mainly the latter group of nuclei; they included correlations *ab initio* by assuming in their calculation a non-spherically symmetric, collective potential in which single-particle states are calculated. Then the coupling of the single-particle motion to the collective motion of the remaining deformed nucleus determines the spectra. (The ingenuity of the Copenhagen concept lies in the clever and successful treatment of the interplay of «collective» and «individual» features of nuclear motion; this provides the model with adequate flexibility to account for all new empirical facts.) It was shown that this easily calculable «unified model», as Aage Bohr likes to call it, could also explain the spectra of nuclei with only a few nucleons outside a closed shell. In this context one should also mention the new work of de-Shalit, in which the first excited states of nuclei with odd A are explained as a combination of «core excitations of the nucleus $A-1$» and the single-particle motion of the odd nucleon.

When one considers all these questions as a whole–the problems of nuclear structure and nuclear forces, as well as the problems of elementary particles–

a verse by Rilke still seems to fairly describe the situation. In the early days of quantum mechanics my late teacher, Wilhelm Lenz, brought this verse to my attention. In it Rilke speaks of his feelings at the turn of the century in terms of a large book in which a page is slowly being turned over, he concludes:

«Man fühlt den Glanz von einer neuen Seite,
Auf der noch alles werden kann.
Die stillen Kräfte prüfen ihre Breite
Und sehn einander dunkel an.»

[*«The lustre of the new-turned page one senses,*
Where everything may yet unfold,
The silent forces measure their expanses;
Each other dimly they behold.»]

Biography

J. Hans D. Jensen was born in Hamburg on 25th June 1907, the son of a gardener Karl Jensen. From 1926 he studied physics, mathematics, physical chemistry and philosophy at the Universities of Hamburg and Freiburg i. Br. He obtained his Ph. D in 1932 in Hamburg (physics, Dr. rer. nat.). He became scientific assistant at the Institute for Theoretical Physics of the University of Hamburg. In 1936 he obtained D. Sc. in Hamburg (Dr. habil.), and he became docent in 1937, and Professor of Theoretical Physics at the Technische Hochschule in Hannover in 1941. In 1949 he was appointed Professor at the University of Heidelberg; since 1969 he is emeritus praecox. In 1947 he was honored with a professorship *h.c.* at the University of Hamburg, and in 1964 with a doctorate *h.c.* at the Technische Universität Hannover. In 1969 he was appointed honorary citizen of Fort Lauderdale (Florida).

He has been a member of the Heidelberg Academy of Sciences since 1947, a corresponding member of the Max Planck Gesellschaft since 1960, and a member of the Sacri Romani Imperii Academia Naturae Coriosorúm (Leopoldina, Halle) since 1964.

He was visiting professor at the University of Wisconsin (1951), the Institute of Advanced Study, Princeton (1952), the University of California at Berkeley (1952), the California Institute of Technology (1953), the Indiana University (1953), the University of Minnesota (1956), and the University of California at La Jolla (1961).

Since 1955 he has been, with O. Haxel, co-editor of the *Zeitschrift für Physik*.

Physics 1964

CHARLES H. TOWNES

NIKOLAI G. BASOV

ALEXANDER M. PROCHOROV

«for fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser–laser principle»

Physics 1964

Presentation Speech by Professor B. Edlén, member of the Nobel Committee for Physics

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

The Nobel prize for physics is in this year given for the invention of the maser and the laser. «Maser» stands for «microwave amplification by stimulated emission of radiation», and the word «laser» is obtained by replacing «microwave» by «light».

The key to the invention is the concept of stimulated emission which was introduced by Einstein already in 1917. By a theoretical analysis of the Planck radiation formula he found that the well-known process of absorption must be accompanied by a complementary process implying that received radiation can stimulate the atoms to emit the same kind of radiation. In this process lies a potential means for amplification. However, the stimulated emission was long regarded as a purely theoretical concept which never could be put to work or even be observed, because the absorption would be the completely dominating process under all normal conditions. An amplification can occur only if the stimulated emission is larger than the absorption, and this in turn requires that there should be more atoms in a high energy state than in a lower one. Such an unstable energy condition in matter is called an inverted population. An essential moment in the invention of the maser and the laser was, therefore, to create an inverted population under such circumstances that the stimulated emission could be used for amplification.

The first papers about the maser were published 10 years ago as a result of investigations carried out simultaneously and independently by Townes and co-workers at Columbia University in New York and by Basov and Prochorov at the Lebedev Institute in Moscow. In the following years there were designed a number of masers of widely different types, and many people made important contributions to this development. In the type that is now being mostly used the maser effect is obtained by means of the ions of certain metals imbedded in a suitable crystal. These masers work as extremely sensitive receivers for short radiowaves. They are of great importance in radio astronomy and are being used in space research for recording the radio signals from satellites.

The optical maser, that is, the laser, dates from 1958, when the possibilities

of applying the maser principle in the optical region were analysed by Schawlow and Townes as well as in the Lebedev Institute. Two years later the first laser was operating.

The step from the microwaves to visible light means a 100 000-fold increase in frequency and causes such changes in the operation conditions that the laser may be regarded as an essentially new invention. In order to achieve the high radiation density required for the stimulated emission to become dominating, the radiating matter is enclosed between two mirrors that force the light to traverse the matter many times. During this process the stimulated radiation grows like an avalanche until all the atoms have given up their energy to the radiation. The fact that the stimulated and stimulating radiation have exactly the same phase and frequency is essential for the result of the process. By virtue of resonance all parts of the active medium combine their forces to give one strong wave. The laser emits what is called coherent light, and this is the decisive difference between the laser and an ordinary light source where the atoms radiate quite independent of each other.

Lasers have now been made in many different shapes. The first, and still most frequently used, type consists of a ruby rod, a few inches long, with the polished and silvered end faces serving as mirrors. The radiation leaves eventually the crystal through one of the end faces which is made slightly transparent. The ruby consists of aluminium oxide with a small admixture of chromium. The chromium ions give to the ruby its red colour, and they are also responsible for the laser effect. The inverted population is produced by the light from a xenon flash lamp. This is absorbed by the ions, putting them in such a condition that they can be stimulated to emit a red light with a well-defined wavelength.

Normally, a large number of successive pulses of laser light is emitted during the time of one flash from the lamp, but by retarding the release until the stored energy has reached a maximum all the energy can be put into one big pulse. The power of the emitted light can then reach more than a hundred million watts. Since, moreover, the emerging ray bundle is strictly parallel, the whole energy can be concentrated by means of a lens on a very small area, producing an enormous power per unit area. From a scientific point of view it is especially interesting that the electrical field strength produced in the light wave may amount to some hundred million volts/cm and thus surpass the forces that keep the electron shells of the atoms together. The high photon density opens up quite new possibilities for studying the interaction of radiation and matter.

Another type of laser, in which the light is emitted from a gas excited by an electric discharge, produces continuously a radiation with a very sharply defined wavelength. This radiation can be used for measurements of lengths and velocities with a previously unattainable precision.

The invention of the laser has provided us with a powerful new tool for research in many fields, the exploitation of which has only just started. Its potential technical applications have been much publicised and are therefore well known. Regarding, especially, the extreme power concentration obtainable with a laser, it should be noted that this effect is limited to short time intervals and very small volumes and therefore attains its main importance for micro-scale operations. It should be emphasized, finally, that the use of a laser beam for destructive purposes over large distances is wholly unrealistic. The «death ray» is and remains a myth.

Dr. Townes, Dr. Basov and Dr. Prochorov. By your ingenious studies of fundamental aspects of the interaction between matter and radiation you have made the atoms work for us in a new and most remarkable way. These magic devices called maser and laser have opened up vast new fields for research and applications which are being exploited with increasing intensity in many laboratories all over the world. On behalf of the Royal Swedish Academy of Sciences I extend to you our warm congratulations and now ask you to receive the Nobel prize from the hands of His Majesty the King.

CHARLES H. TOWNES

Production of coherent radiation by atoms and molecules

Nobel Lecture, December 11, 1964

From the time when man first saw the sunlight until very recently, the light which he has used has come dominantly from spontaneous emission, like the random emission of incandescent sources. So have most other types of electromagnetic radiation–infrared, ultraviolet, or gamma rays. The maximum radiation intensities, or specifically the power radiated per unit area per unit solid angle per unit frequency bandwidth, have been controlled by Planck's black-body law for radiation from hot objects. This sets an upper limit on radiation intensity–a limit which increases with increasing temperature, but we have had available temperatures of only a few tens of thousands or possibly a few millions of degrees.

Radio waves have been different. And, perhaps without our realizing it, even much of our thinking about radio waves has been different, in spite of Maxwell's demonstration before their discovery that the equations governing radio waves are identical with those for light. The black-body law made radio waves so weak that emission from hot objects could not, for a long time, have been even detected. Hence their discovery by Hertz and the great use of radio waves depended on the availability of quite different types of sources–oscillators and amplifiers for which the idea of temperature and black-body radiation even seems rather out of place. For example, if we express the radiation intensity of a modern electronic oscillator in terms of temperature, it will typically be in the range 10^{10} to 10^{30} degrees Kelvin.

These two regimes, radio electronics and optics, have now come much closer together in the field known as quantum electronics, and have lent each other interesting insights and powerful techniques.

The development of radar stimulated many important applications of electronics to scientific problems, and what occupied me in particular during the late 1940's was microwave spectroscopy, the study of interactions between microwaves and molecules. From this research, considerable information could be obtained about molecular, atomic, and nuclear structure. For its success, coherent microwave oscillators were crucial in allowing a powerful

high-resolution technique. Consequently it was important for spectroscopy, as well as for some other purposes, to extend their range of operation to wavelengths shorter than the known limit of electronic oscillators, which was near 1 millimeter. Harmonic generation and some special techniques allowed interesting, though rather slow, progress. The basic problem with electronic amplifiers or oscillators seemed to be that inevitably some part of the device which required careful and controlled construction had to be about as small as the wavelength generated. This set a limit to construction of operable devices[1]. It was this experimental difficulty which seemed inevitably to separate the techniques which were applicable in the radio region from those applicable to the shorter waves of infrared or optical radiation.

Why not use the atomic and molecular oscillators already built for us by nature? This had been one recurring theme which was repeatedly rejected. Thermodynamic arguments tell us, in addition to the black-body law of radiation, that the interaction between electromagnetic waves and matter at any temperature* cannot produce amplification, for radiation at the temperature of matter cannot be made more intense by interaction of the two without violating the second law. But already by 1917, Einstein had followed thermodynamic arguments further to examine in some detail the nature of interactions between electromagnetic waves and a quantum-mechanical system. And a review of his conclusions almost immediately suggests a way in which atoms or molecules can in fact amplify.

The rate of change of electromagnetic energy confined in a region where it interacts with a group of molecules must, from Einstein's work, have the form

$$\frac{dI}{dt} = AN_b - BIN_a + B'IN_b \qquad (1)$$

where N_a and N_b are the numbers of molecules in the upper and lower of two quantum states, which we assume for simplicity to be nondegenerate (that is, single). A and B are constants, and thus the first and second terms represent spontaneous emission and absorption, respectively. The third term represents emission from the upper state produced by the presence of a radiation intensity I, and is hence called stimulated emission.

At equilibrium, when

$$\frac{dI}{dt} = 0, \quad I = \frac{AN_b}{BN_a - B'N_b}$$

* Strictly speaking, at any *positive* temperature. Negative absolute temperatures can be defined as will be noted below.

Rather simple further thermodynamic reasoning shows that $B' = B$ and gives the ratio A/B. While Boltzmann's law $N_b = N_a e^{-W/kT}$ requires $N_b < N_a$ at any temperature T, it is immediately clear from Eqn. 1 that if $N_b > N_a$, dI/dt will always be positive and thus the radiation amplified. This condition is of course one of nonequilibrium for the group of molecules, and it hence successfully obviates the limits set by black-body radiation. The condition $N_b > N_a$ is also sometimes described as population inversion, or as a negative temperature[2], since in Boltzmann's law it may be obtained by assuming a negative absolute temperature.

Thermodynamic equilibrium between two states of a group of atoms requires not only a Boltzmann relation $N_b = N_a e^{-W/kT}$ but also a randomness of phases of the wave functions for the atoms. In classical terms, this means that, if the atomic electrons are oscillating in each atom, there must not be a correlation in their phases if the entire group can be described as in temperature equilibrium. Einstein's relation (Eqn. 1) in fact assumed that the phases are random. And, if they are not, we have another condition which will allow the atoms to amplify electromagnetic waves, even when $N_b < N_a$. This represents a second type of loophole in the limits set by the black-body law and thermodynamic equilibrium, and one which can also be used alone or in conjunction with the first in order to produce amplification.

Thermodynamic arguments can be pushed further to show that stimulated emission (or absorption) is coherent with the stimulating radiation. That is, the energy delivered by the molecular systems has the same field distribution and frequency as the stimulating radiation and hence a constant (possibly zero) phase difference. This can also be shown somewhat more explicitly by a quantum-mechanical calculation of the transition process.

Stimulated emission received little attention from experimentalists during the 1920's and 1930's when atomic and molecular spectroscopy were of central interest to many physicists.

Later, in the 1940's, experiments to demonstrate stimulated emission were at least discussed informally and were on the minds of several radio spectroscopists, including myself. But they seemed only rather difficult demonstrations and not quite worth while. In the beautiful 1950 paper of Lamb and Retherford on the fine structure of hydrogen[3] there is a specific brief note about «negative absorption» with reversal of population. And a year later Purcell and Pound[4] published their striking demonstration of population inversion and stimulated emission. As a matter of fact, population inversion and its effects on radiation had already shown up in a somewhat less accented form in

the resonance experiments of Bloch[5] and others. But all these effects were so small that any amplification was swamped by losses due to other competing processes, and their use for amplification seems not to have been seriously considered until the work of Basov and Prokhorov[6], Weber[7], and of Gordon, Zeiger, and Townes[8,9] in the early 1950's.

My own particular interest came about from the realization that probably only through the use of molecular or atomic resonances could coherent oscillators for very short waves be made, and the sudden discovery in 1951 of a particular scheme[8] which seemed to really offer the possibility of substantial generation of short waves by molecular amplification.

Basic Maser Principles

The crucial requirement for generation, which was also recognized by Basov and Prokhorov, was to produce positive feedback by some resonant circuit and to ensure that the gain in energy afforded the wave by stimulated molecular transitions was greater than the circuit losses. Consider a resonant microwave cavity with conducting walls, a volume V, and a quality factor Q. The latter is defined by the fact that power lost because of resistance in the walls is

$$\frac{\overline{E}^2 V\nu}{4Q}$$

where $\overline{E}^2$ is the electric field strength in the mode averaged over the volume and ν is the frequency. If a molecule in an excited state is placed in a particular field of strength E, the rate of transfer of energy to the field is

$$\left(\frac{\overline{E}\mu}{\hbar}\right)^2 \frac{h\nu}{3\Delta\nu}$$

when the field's frequency coincides with the resonance frequency ν between the two molecular states. Here μ is a dipole matrix element for the molecular transition and $\Delta\nu$ is the width of the molecular resonance at half maximum (if a Lorentz line shape is assumed). Hence for N_b molecules in the upper state and N_a in the lower state the power given the field in the cavity is

$$(N_b - N_a)\left(\frac{\overline{E}\mu}{\hbar}\right)^2 \frac{h\nu}{3\Delta\nu}$$

If the molecules are distributed uniformly throughout the cavity, E^2 must be averaged over the volume. For the net power gain to be positive, then,

$$(N_b - N_a)\left(\frac{\bar{E}\mu}{\hbar}\right)^2 \frac{h\nu}{3\Delta\nu} \geq \frac{\bar{E}^2 V\nu}{4Q}$$

This gives the threshold condition for buildup of oscillations in the cavity

$$(N_b - N_a) \geq \frac{3h\, V\Delta\nu}{16\pi^2 Q\mu^2} \qquad (2)$$

There is by now an enormous variety of ways in which the threshold condition can be met, and some of them are strikingly simple. But the system which first seemed to give an immediate hope of such an oscillator involved a beam of ammonia molecules entering a resonant cavity, as shown in Fig. 1. The

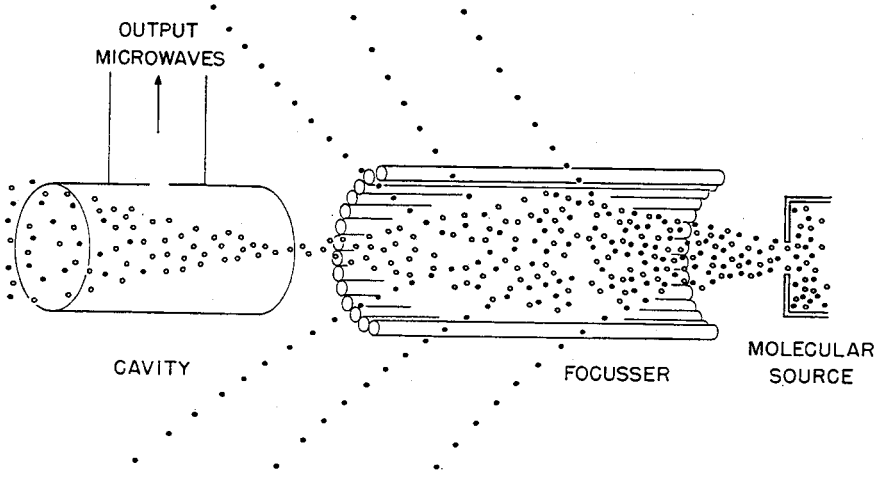


Fig. 1. The ammonia (beam-type) maser. Molecules diffuse from the source into a focusser where the excited molecules (open circles) are focused into a cavity and molecules in the ground state (solid circles) are rejected. A sufficient number of excited molecules will initiate an oscillating electromagnetic field in the cavity, which is emitted as the output microwaves. Because of energy given to the field, some molecules return to the ground state toward the end of their transit through the cavity.

transition used was the well-known inversion transition of ammonia at 23,870 Mc/sec. A «focuser», involving inhomogeneous electric fields, tends to remove molecules in the ground state from the beam and to focus molecules in the excited state along the axis of the beam and into the cavity, thus ensuring

that $N_b > > N_a$. J. P. Gordon played a crucial role in making operable the first such system in 1954, after 2½ years of experimental work[9,10], and H. J. Zeiger was a valuable colleague in the first year of work and early designs. We called this general type of system the maser, an acronym for *m*icrowave *a*mplication by *s*timulated *e*mission of *r*adiation. The idea has been successfully extended to such a variety of devices and frequencies that it is probably well to generalize the name—perhaps to mean *m*olecular *a*mplification by *s*timulated *e*mission of *r*adiation. But in the radio-frequency range it is sometimes called the *r*aser, and for light the term *l*aser is convenient and commonly used. Maser amplification is the key process in the new field known as quantum electronics—that is, electronics in which phenomena of a specifically quantum-mechanical nature play a prominent role.

It is well known that an amplifier can usually be made into an oscillator, or *vice versa*, with relatively minor modifications. But it was only after experimental work on the maser was started that we realized this type of amplifier is exceedingly noise-free. The general reason for low noise can be stated simply. The molecules themselves are uncharged so that their motions, in contrast to motions of electrons through vacuum tube amplifiers, produce no unwanted electromagnetic signals. Hence a signal introduced into the resonant cavity competes only with whatever thermal noise is in the cavity as the result of thermal radiation from the cavity walls, and with spontaneous emission from the excited molecules. Spontaneous emission can be regarded for this purpose as that stimulated by a fluctuating field of energy $h\nu$. Since $kT \approx 200 h\nu$ for microwaves in a cavity at room temperature, the thermal radiation kT in the cavity is much more important than spontaneous emission. It is then only the thermal radiation present which sets the limit to background noise, since it is amplified precisely as is the signal.

The above discussion also shows that, if the cavity is at 0°K and no extraneous noise enters the cavity with the input signal, the limiting noise fluctuation is determined by the spontaneous emission, which is equivalent to only one quantum of energy in the cavity. It can be shown, in fact, that masers can yield the most perfect amplification allowed by the uncertainty principle.

The motion of an electromagnetic wave is analogous to that of a mechanical harmonic oscillator, the electric and magnetic fields corresponding to position and momentum of the oscillator. Hence the quantum-mechanical uncertainty principle produces an uncertainty in the simultaneous determination of the electric and magnetic fields in a wave, or equivalently in determination of the total energy and phase of the wave. Thus one can show that, to

the extent that phase of an electromagnetic wave can be defined by a quantum-mechanical operator, there is an uncertainty relation[11]

$$\Delta n \Delta \phi \geq {}^1/_2 \tag{3}$$

Here Δn is the uncertainty in the number of photons in the wave, and $\Delta \phi$ is the uncertainty in phase measured in radians.

Any amplifier which gives some representation of the phase and energy of an input wave in its output must, then, necessarily involve uncertainties or fluctuations in intensity. Consider, for example, an ideal maser amplifier composed of a large number of molecules in the upper state interacting with an initial electromagnetic wave, which is considered the signal. After some period of time, the electromagnetic wave will have grown to such magnitude that it contains a very large number of quanta and hence its phase and energy can be measured by classical means. By using the expected or average gain and phase relation between the final electromagnetic wave and the initial signal the maser amplifier thus allows a measurement of the initial wave.

A calculation by well-established quantum-mechanical techniques of the relation between input and output waves shows that this measurement of the input wave leaves an uncertainty just equal to the minimum required by the uncertainty principle[11]. Furthermore, the product $\Delta E \Delta H$ of uncertainties in the electric and magnetic fields has the minimum value allowed while at the same time $(\Delta E)^2 + (\Delta H)^2$ is minimized. The uncertainty in number n of quanta in the initial wave is

$$\Delta n = \sqrt{n+1}$$

and in phase it is

$$\Delta \phi = \frac{1}{2\sqrt{n}}$$

so that

$$\Delta n \Delta \phi = {}^1/_2 \frac{\sqrt{n+1}}{n}$$

The phase has real meaning, however, only when there are as many as several quanta, in which case $\Delta n \Delta \phi \to \frac{1}{2}$, the minimum allowed by Eqn. 3. The background noise, which is present with no input signal ($n = 0$), is seen to be equivalent to a single quantum ($\Delta n = 1$) of input signal.

A somewhat less ideal maser might be made of N_a and N_b molecules in the upper and lower states, respectively, all interacting with the input signal. In this case fluctuations are increased by the ratio $N_b/(N_b - N_a)$. If the amplifier

has a continuous input signal, a continuous amplified output, and a bandwidth for amplification $\Delta\nu$, the noise power output can be shown to be equivalent to that produced by an input signal[12]

$$N = \frac{h\nu\Delta\nu}{1 - \frac{N_a}{N_b}}$$

The noise power N is customarily described in terms of the noise temperature T_n of the amplifier, defined by $N = kT_n\Delta\nu$. Thus the minimum noise temperature allowed by quantum mechanics is that for a maser with $(N_a/N_b) << 1$, which is

$$T_n = \frac{h\nu}{k} \tag{4}$$

This is equivalent to the minimum energy uncertainty indicated above of one quantum ($\Delta n = 1$). In the microwave region, T_n given by Eqn. 4 is approximately 1°, whereas the best other microwave amplifiers when maser amplifiers were first being developed had noise fluctuations about 1 000 times greater.

It is interesting to compare an ideal maser as a detector with a perfect photodetector, such as a γ-ray counter. The γ-ray counter can detect a single photon with almost no false signals, whereas a maser must always have a possible false signal of about one photon. But the photodetector gives no information about the phase of the signal; it only counts quanta, which is why the uncertainty principle allows $\Delta n \to 0$. Unfortunately, there are no perfect photodetectors in the microwave or radio regions, so that the maser is our best available detector for these waves.

The same freedom from noise which makes the maser a good amplifier helps make it a strikingly good source of monochromatic radiation since, when the threshold condition is fulfilled and the maser oscillates, the low noise implies a minimum of random frequency fluctuations.

Consider now a maser oscillator consisting of a group of excited molecules in a resonant cavity. Let the molecular transition frequency be ν_m, its half width at half-maximum intensity $\Delta\nu_m$, and the resonant-cavity frequency be ν_c with a half width $\Delta\nu_c$. If ν_m and ν_c differ by much less than $\Delta\nu_m + \Delta\nu_c$, the radiation produced by the oscillation can be shown to occur at a frequency[13]

$$\nu = \frac{\nu_m Q_m + \nu_c Q_c}{Q_m + Q_c} \tag{5}$$

where the quality factors Q_m and Q_c are $\nu_m/\Delta\nu_c$ and $\nu_m/\Delta\nu_c$ respectively. Thus if the molecular resonance is much sharper than that of the cavity, as in the ammonia-beam maser ($Q_m >> Q_c$), the frequency of oscillation is[10]

$$\nu = \nu_m + (\nu_c - \nu_m)\frac{Q_c}{Q_m} \tag{6}$$

If the cavity is tuned so that $\nu_c - \nu_m$ is small, then the frequency of oscillation coincides very closely with the natural molecular frequency ν_m, and one has an almost constant frequency oscillator based on a molecular motion, a so-called atomic clock.

The frequency ν is not precisely defined or measurable because of noise fluctuations, which produce random phase fluctuations of the wave. In fact, the maser is essentially like a positive feedback amplifier which amplifies whatever noise source happens to be present and thereby produces a more or less steady oscillation. If Q_m or Q_c is high, and the amplifier gain is very large, then the bandwidth of the system becomes exceedingly small. But it is never zero, nor is the frequency ever precisely defined. The average deviation in frequency from Eqn. 5 which these phase fluctuations produce when averaged over a time t is[14]

$$\varepsilon = \Delta\nu\left(\frac{W_n}{Pt}\right)^{\frac{1}{2}} \tag{7}$$

where

$$\Delta\nu = \frac{\Delta\nu_c\,\Delta\nu_m}{\Delta\nu_c + \Delta\nu_m}$$

P is the power generated by the oscillator, and W_n is the effective energy in the source of fluctuations. Where $kT >> h\nu$ in a cavity at temperature T and resonant frequency ν, the effective energy comes from thermal noise and $W_n = kT$. If the noise fluctuations come from spontaneous emission, as they do when $kT << h\nu$, then $W_n = h\nu$.

It is also useful to state the spectral width of the radiation emitted from a maser oscillator, as well as the precision to which the frequency can be determined. The half width of the spectral distribution is again determined by the same noise fluctuation and is given by[10,15,16]

$$\delta = \frac{2\pi W_n}{P}(\Delta\nu)^2 \tag{8}$$

where $\Delta\nu$, W_n, and P are the same as in Eqn. 7. This width δ is typically so small in maser oscillators that they provide by far the most monochromatic sources of radiation available at their frequencies.

Maser Clocks and Amplifiers

Although the ammonia beam-type maser was able to demonstrate the low-noise amplification which was predicted[17], its extremely narrow band-width makes it and other beam-type masers more useful as a very monochromatic source of electromagnetic waves than as an amplifier. For the original maser, the power output P was about 10^{-9} Watt, and the resonance width $\Delta\nu$ about 2 kilocycles, as determined by the length of time required for the beam of molecules to pass through the cavity. Since the frequency of oscillation ν_m is 23,874 megacylces, the fractional spectral width, according to Eqn. 8, is $\delta/\nu \approx 10^{-14}$. In a time $t = 100$ seconds, Eqn. 7 shows that the frequency can be specified to a fractional precision $\varepsilon/\nu = 2 \cdot 10^{-14}$, and of course the precision increases for longer times proportionally to $1/t^{1/2}$.

As a constant-frequency oscillator or precise atomic clock, however, the ammonia maser has an additional problem which is not so fundamental, but which sets a limit on long-term stability. This comes from long-term drifts, particularly of the cavity temperature, which vary ν_c. These variations can be seen, from Eqn. 6, to «pull» ν. Variations of this type have limited the long-term stability[18] of ammonia masers to fractional variations of about 10^{-11}; this still represents a remarkably good clock.

A beam-type maser using the hyperfine structure transition in the ground state of hydrogen, which is at 1420 megacycles, has recently been developed by Goldenberg, Kleppner, and Ramsey[19]. In this case, the excited atoms bounce many times from glass walls in the cavity, and thereby a resonance width as small as 1 cycle per second is achieved. Present designs of the hydrogen maser yield an oscillator with long-term fractional variations no larger than about 10^{-13}. This system seems likely to produce our best available clock or time standard.

Masers of reasonably wide utility as amplifiers came into view with the realization that certain solids containing paramagnetic impurities allowed attainment of the maser threshold condition[20]. Microwave resonances of paramagnetic atoms in solids, or in liquids, had been studied for some time, and many of their properties were already well known. The widths of these re-

sonances vary with materials and with impurity concentration from a small fraction of a megacycle to many hundreds of megacycles, and their frequencies depend on applied magnetic field strengths, so that they are easily tunable. Thus they offer for maser amplifiers a choice of a considerable range of bandwidth, and a continuous range of frequencies.

A paramagnetic atom of spin 1/2 has two energy levels which, when placed in a magnetic field, are separated by an amount usually of about $\nu = 2.8\,H$ Mc. Here H is the field in gauss, and from this it is clear that most of the microwave frequency range can be covered by magnetic fields of normal magnitudes. The first paramagnetic masers suggested involved impurity atoms of this type in crystals of silicon or germanium. Relaxation between the two states was slow enough in these cases that a sufficient population inversion could be achieved[20]. However, before very long a very much more convenient scheme for using paramagnetic resonances was proposed by Bloembergen[21], the so-called three-level solid-state maser. This system allowed continuous inversion of population, and hence continuous amplification, which was very awkward to obtain in the previous two-level system.

Paramagnetic atoms with an angular momentum due to electron spin S greater than 1/2 have $2S+1$ levels which are degenerate when the atom is in free space. But these levels may be split by «crystalline fields», or interaction

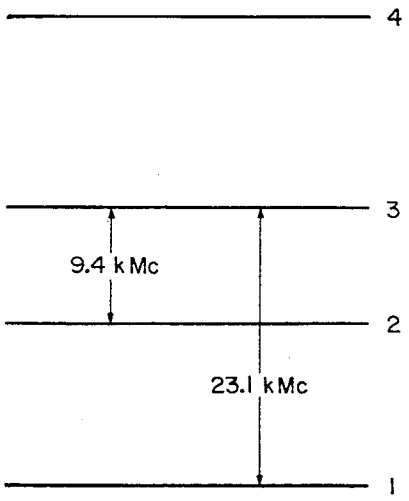


Fig. 2. Energy levels of Cr^{2+} in ruby with a particular crystalline orientation in a magnetic field of 3900 oersteds. For a three-level maser, 23.1 kMc ($23.1 \cdot 10^3$ Mc) is the frequency of the pumping field and 9.4 kMc is the frequency of amplification or oscillation.

with neighboring atoms if the atoms are imbedded in a solid, and frequently the splittings lie in the microwave range. The energy levels of such a system, involving a spin of 3/2 and four levels, can be as indicated in Fig. 2 when the system is in a magnetic field. If a sufficiently large electromagnetic wave of frequency ν_{13} (the transition frequency between levels 1 and 3) is applied, the population of these two levels can be equalized or «saturated». In this case, the ratio of the population of level 2 to that of level 1 or 3 under steady conditions is

$$\frac{n_2}{n_1} = \frac{\frac{1}{T_{12}} e^{\frac{-h\nu_{12}}{kT}} + \frac{1}{T_{23}}}{\frac{1}{T_{12}} + \frac{1}{T_{23}} e^{\frac{-h\nu_{23}}{kT}}}$$

Here T is the temperature of the crystal containing the impurities, and T_{12} and T_{23} are the times for relaxation between the states 1 and 2 or 2 and 3, respectively. For $h\nu_{12} > > kT$ and $h\nu_{23} > > kT$, as occurs at very low temperatures or at ordinary temperatures if the levels are separated by optical frequencies,

$$\frac{n_2}{n_1} = \frac{T_{12}}{T_{23}}$$

When $h\nu_{12} < < kT$ and $h\nu_{23} < < kT$, which is more commonly the case for microwaves,

$$\frac{n_2}{n_1} = 1 + \frac{h}{kT} \frac{\frac{\nu_{12}}{T_{12}} - \frac{\nu_{23}}{T_{23}}}{\frac{1}{T_{12}} + \frac{1}{T_{23}}} \tag{9}$$

Thus if

$$\frac{\nu_{12}}{T_{12}} > \frac{\nu_{23}}{T_{23}}$$

there is an inversion of population between levels 2 and 1, or if,

$$\frac{\nu_{12}}{T_{12}} < \frac{\nu_{23}}{T_{23}}$$

there is an inversion of population between levels 3 and 2, since the populations n_3 and n have been equalized by the «pumping» radiation. Equation 9 is essentially the result obtained by Bloembergen[21], who also suggested several promising paramagnetic materials which might be used. Basov and Prok-

horov had already proposed a rather similar three-level «pumping» scheme for application to a molecular beam system[22].

The first successful paramagnetic maser of this general type was obtained by Scovil *et al.*[23], using a rare-earth ion in a water-soluble crystal. But, before long, other more suitable crystals such as ruby[24] (chromium ions in Al_2O_3) became more or less standard and have provided amplifiers of remarkable sensitivity for radio astronomy, for satellite communication, and for communication with space probes[25]. They have considerably improved the potentialities of radio astronomy, and have already led to some new discoveries[26, 27]. These systems generally require cooling with liquid helium, which is a technological difficulty that some day may be obviated. But otherwise they represent rather serviceable and convenient amplifiers.

A maser amplifier of microwaves can rather easily be built which has a theoretical noise temperature as low as 1° or 2°K, and experimental measurements have confirmed this figure[28]. However, such a low noise level is not easy to measure because almost any measurement involves attachment of input and output circuits which are at temperatures much higher than 1°K, and which radiate some additional noise into the amplifier. The lowest overall noise temperature so far reported for an entire receiving system[29] using a maser amplifier is about 10°K. This represents about 100 times the sensitivity of microwave amplifiers built before invention of the maser. But masers have stimulated other amplifier work, and some parametric amplifiers, using more or less classical properties of materials rather than quantum electronics, now have sensitivities within a factor of about 5 of this figure.

Optical and Infrared Masers, or Lasers

Until about 1957, the coherent generation of frequencies higher than those which could be obtained from electronic oscillators still had not been directly attacked, although several schemes using molecular-beam masers for the far-infrared were examined from time to time. This lack of attention to what had been an original goal of the maser came about partly because the preliminary stages, including microwave oscillators, low-noise amplifiers, and their use in various scientific experiments, had proved so interesting that they distracted attention from the high-frequency possibilities.

But joint work with A. L. Schawlow[30], beginning at about this time, helped open the way for fairly rapid and interesting development of maser oscillators

in the far-infrared, optical, and ultraviolet regions – as much as 1000 times higher in frequency than any coherent sources of radiation previously available. It is masers in these regions of the spectrum, frequently called lasers (*l*ight *a*mplification by *s*timulated *e*mission of *r*adiation), which have perhaps provided the most striking new scientific tools and results. Important aspects of this work were clear demonstrations that there are practical systems which can meet the threshold condition of oscillation, and that particular resonator designs allow the oscillations to be confined to certain specific and desirable modes. The resonator analyzed was composed simply of two parallel mirrors – the well-known Fabry–Perot interferometer, but of special dimensions.

For light waves, the wavelength is so short that any macroscopic resonator constructed must have dimensions that are large compared with the wavelength. In this case, the electromagnetic field may to some reasonable approximation be considered to travel in straight lines and be reflected from the walls of the resonator. The threshold condition may be written

$$\left(\frac{\mu E}{\hbar}\right)^2 \frac{h\nu\,(N_b - N_a)}{12\pi\,\Delta\nu} \geq \frac{E^2}{8\pi}\,\frac{V}{t} \tag{10}$$

where t is the decay time for the light in a cavity of reflecting walls and volume V. If the light has a random path in the cavity, the decay time can be expressed generally in terms of the reflection coefficient r of the walls, the volume V, the wall area A, and the velocity of light c,

$$t = \frac{6V}{(1-r)Ac}$$

Hence Eqn. 10 becomes[30]

$$N_b - N_a \geq \frac{\Delta\nu}{\nu}\,\frac{h\,(1-r)Ac}{16\pi^2\,\mu^2\,V} \tag{11}$$

It can be seen that this critical condition is almost independent of frequency if the fractional line width $\Delta\nu/\nu$ does not change with frequency (as, for example, in the Doppler effect). The reflection coefficient and dipole moment matrix element μ are not particularly dependent on frequency over the range in question. Hence, if the critical condition can be met for one frequency, it can probably be met over the entire range from the far-infrared to the ultraviolet.

There is a problem with a resonator which is large compared to a wavelength in that there are many modes. Hence, unless the modes in which oscilla-

tions occur are successfully controlled, the electromagnetic field may build up simultaneously in many modes and at many frequencies. The total number of modes in a cavity with frequencies which lie within the line width $\Delta\nu$ of the atomic molecular resonance is

$$p = \frac{8\pi^2 V \nu^2 \Delta\nu}{c^3}$$

or about 10^9 for a cavity volume of 1 cm^3, a frequency in the optical region, and ordinary atomic line widths. But fortunately the possibility of oscillation can be eliminated for most of these modes.

Two small parallel mirrors separated by a distance much larger than their diameter will allow a beam of light traveling along the axis joining them to travel back and forth many times. For such a beam, the decay time t is $L/c(1-r)$, where L is the mirror separation and r the reflectivity. Hence the threshold condition is

$$N_b - N_a \geq \frac{3\Delta\nu}{8\pi^2\nu} \frac{hc(1-r)}{\mu^2 L}$$

This assumes that diffraction losses are negligible. A beam of light which is not traveling in a direction parallel to the axis will disappear from the volume between the mirrors much more rapidly. Hence the threshold condition for off-axis beams will require appreciably more excited atoms than that for axial beams, and the condition for oscillation can be met for the latter without a build-up of energy in off-axis light waves.

Many features of the modes for the electromagnetic wave between two square, plane, parallel mirrors of dimension D and separation L can be approximately described as those in a rectangular box of these dimensions, although the boundary conditions on the enclosed sides of the «box» are of course somewhat different. The resonant wavelengths of such a region for waves traveling back and forth in a nearly axial direction are[30]

$$\lambda = \frac{2L}{q}\left[1 - \tfrac{1}{2}\left(\frac{Lr}{Dq}\right)^2 - \tfrac{1}{2}\left(\frac{Ls}{Dq}\right)^2\right] \qquad (12)$$

where q, r, and s are integers, and $r << q$, $s << q$. More precise examination of the modes requires detailed numerical calculation[31]. For a precisely axial direction, $r = s = 0$, and the modes are separated by a frequency $c/2L$. If this frequency is somewhat greater than the atomic line width $\Delta\nu$, then only one axial mode can oscillate at a time. The axial wave has an angular width due to

diffraction of about λ/D, and, if this is comparable with the angle D/L, then all off-axis modes (r or $s \neq 0$) are appreciably more lossy than are the axial ones, and their oscillations are suppressed.

If one of the mirrors is partially transparent, some of the light escapes from the axial mode in an approximately plane wave and with an angular divergence, approximately λD, determined by diffraction.

A number of modified resonator designs have been popular and useful in optical masers, in particular ones based on the confocal Fabry–Perot interferometer. However, the plane-parallel case seems to offer the simplest means of selecting an individual mode.

Although a number of types of atomic systems and excitation seemed promising in 1958 as bases for optical masers, optical excitation of the alkali vapors lent itself to the most complete analysis and planning for an operable oscillator. One such system has been shown to oscillate as expected[32]; but the alkali vapors are no longer of great interest, because other systems which were at the time much less predictable have turned out to be considerably more useful.

The first operating laser, a system involving optical excitation of the chromium ions in ruby and yielding red light, was demonstrated by Maiman in 1960[33]. He took what seemed at first a rather difficult route of inverting the population between the ground state and excited states of the chromium ion. This technique requires that at least half of the very large number of atoms in the ground state must be excited in order to have the possibility of a population inversion. In the case of two normally unpopulated atomic states, the total amount of excitation required is much less. However, Maiman succeeded handsomely in exciting more than half the chromium ions in a ruby with chromium concentration of about 1/2000 by applying a very intense pulse of light from a flash tube. This type of system is illustrated schematically in Fig. 3. Success immediately yielded a very-high-energy maser oscillation because, to get population inversion at all, a large amount of energy must be stored in the excited atomic states. Surfaces of the ruby served as the reflecting mirrors. Collins *et al.*[34] quickly demonstrated that the ruby laser showed many of the characteristics predicted for such an oscillator.

The ruby laser is operated normally only in pulses, because of the high power required to reach threshold, and emits intense bursts of red light at power levels between about 1 kilowatt and 100 megawatts. It has given rise to a whole family of lasers involving impurities in various crystals of glasses, and covering frequencies from the near infrared into the optical region.

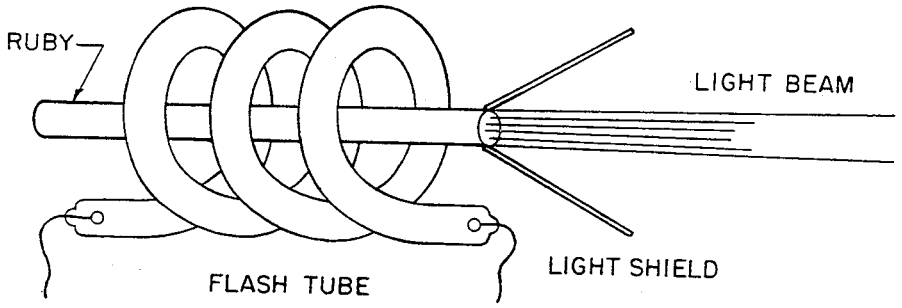


Fig. 3. Schematic diagram of a ruby (optically excited solid-state) laser. When the gas flash tube is activated, electromagnetic oscillations occur within the ruby rod, and some of these light waves are emitted in a beam through one partially reflecting end of the rod.

Not very long after the ruby laser was developed, Javan, Bennett and Herriott[35] obtained maser oscillations from neon atoms excited by collisions of the second kind with metastable helium, in accordance with an idea previously put forward by Javan[36]. This system, illustrated in Fig. 4, requires only a gaseous discharge through a tube containing a mixture of helium and neon at low pressure, and two reflectors at the ends of the tube. It oscillates at the

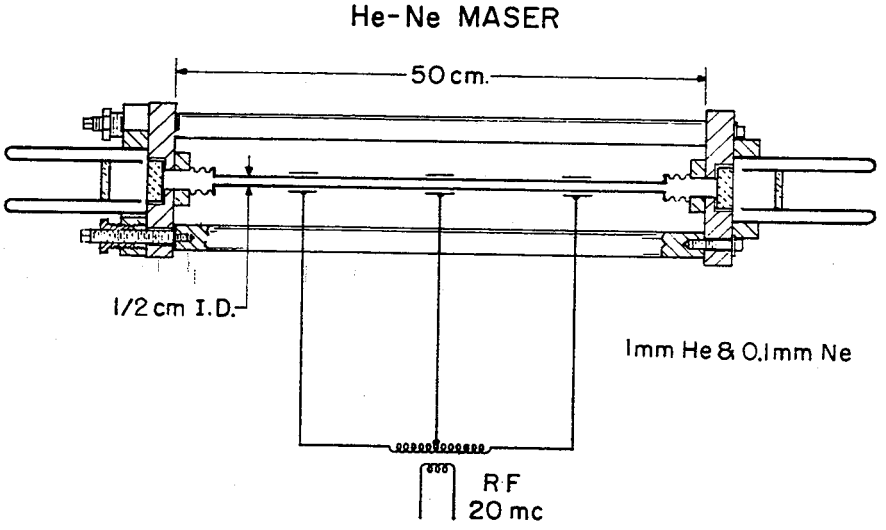


Fig. 4. Schematic diagram of a helium–neon (gas discharge) laser. Electrical excitation can initiate a steady maser oscillation, resulting in an emitted light beam from either end of the gas discharge, where there are reflecting mirrors.

relatively low power of about 1 milliwatt, but approaches ideal conditions much more closely than the ruby system, and affords a continuous source of infrared radiation of great purity and directivity.

The technique of gaseous excitation by electrical discharge has also led to a large family of lasers, producing hundreds of different frequencies from many different gases which range from wavelengths as long as a few tenths of 1 millimeter down into the ultraviolet. For some systems, a heavy discharge pulse in the gas is needed. Others, particularly some of the infrared frequencies in rare gases, oscillate so readily that it seems probable that we have had lasers accidentally all along. Very likely some neon or other rare-gas electric signs have been producing maser oscillations at infrared wavelengths, which have gone unnoticed because the infrared could not escape from the glass neon tubes. Some of these oscillation frequencies represent atomic transitions which were previously undetected; for others, the transition has not yet even been identified.

Another class of lasers was initiated through the discovery[37] that a *p–n* junction of the semiconductor gallium arsenide through which a current is passed can emit near-infrared light from recombination processes with very high efficiency. Hall *et al.*[38] obtained the first maser oscillations with such a system, with light traveling parallel to the junction and reflected back and forth between the faces of the small gallium arsenide crystal. His results were paralleled or followed immediately, however, by similar work in two other laboratories[39, 40]. This type of laser, illustrated in Fig. 5, is of the general size and cost of a transistor. It can be made to oscillate simply by passage of an electric current, and in some cases the radiation emitted represents more than 50 percent of the input electrical energy – an efficiency greater than that of other man-made light sources.

There quickly developed a large family of semiconductor lasers, some involving junctions and, recently, some using excitation by an external beam of electrons[41]. They range in wavelength from about 10 microns, in the infrared, to the center of the visible region.

Normal Raman scattering can be regarded as spontaneous emission from a virtual state, as indicated in Fig. 6. Associated with any such spontaneous emission there must be, in accordance with Einstein's relations, a stimulated emission. Javan showed[42] the principles involved in using this stimulated emission for a Raman maser. What is required is simply a large enough number of molecular systems which are sufficiently strongly excited by radiation of frequency greater than some Raman-allowed transition.

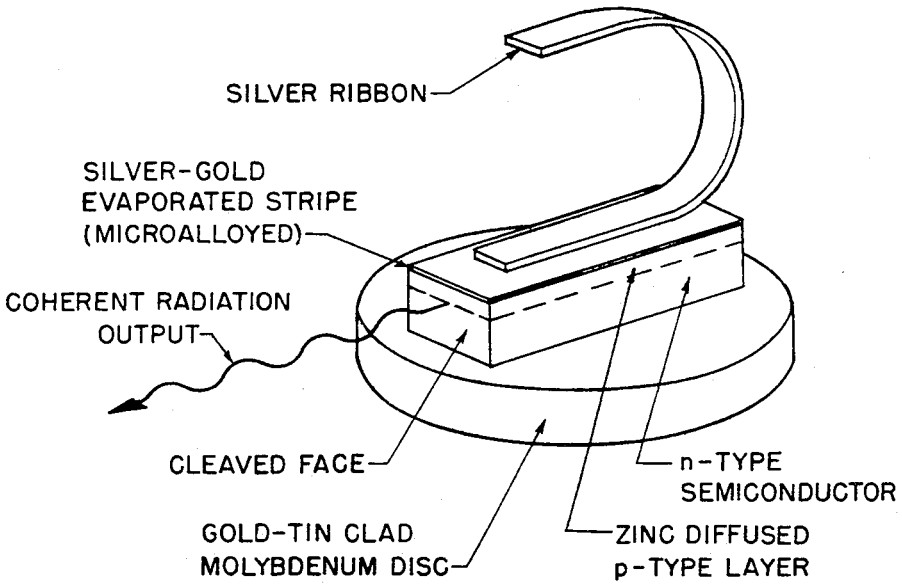


Fig. 5. Schematic diagram of a gallium arsenide (injection, or semiconductor) laser. A small voltage applied between the silver ribbon and the molybdenum disc can produce maser oscillations with resulting emission of coherent infrared radiation.

One might consider the population of the virtual level in a Raman maser (see Fig. 6) to be greater than that of the first excited state, so that there is no population inversion. On the other hand the initial state, which is the ground state, needs to be more populated than the first excited state. One can quite properly consider the amplification process as a parametric one with the molecular frequency as idler, or as due to a mixture of ground and excited states in which there is phase coherence between the various molecules. This is the second type of loophole through the black-body radiation law mentioned earlier. The ammonia-beam maser itself illustrates the case of amplification without the necessity of population inversion. As the ammonia molecules progress through the cavity and become predominantly in the ground state rather than the excited state, they continue to amplify because their oscillations are correlated in phase with each other, and have the appropriate phase with respect to the electromagnetic wave.

Raman masers were first demonstrated by Woodbury and Ng[43] as the result of excitation of various liquid molecules with a very intense beam from a pulsed laser. They too have now many versions, giving frequencies which

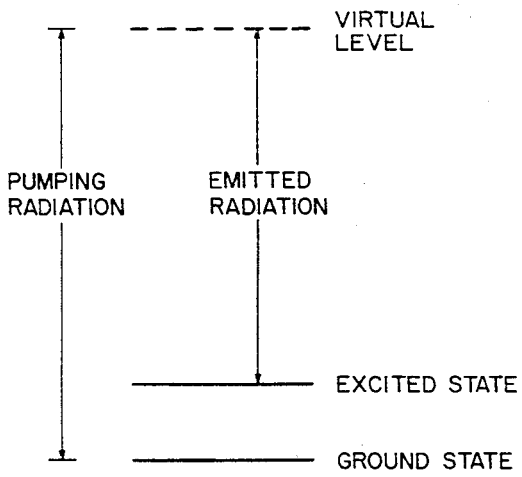


Fig. 6. Representation of energy levels in a Raman maser. This system resembles qualitatively a three-level maser, one of the levels being «virtually», or not characteristic of the molecule when no field is present.

differ from the original driving maser beam by some small integer times a molecular-vibrational frequency. Their action has been considerably extended by Terhune[44] and treated in a number of theoretical papers[42,45].

Present Performance of Lasers

Where now do we stand in achieving the various theoretical expectations for performance of masers?

First, consider the general extension of the frequency range where we have coherent amplifiers and oscillators. This has been increased by a factor of somewhat more than 1 000; there are still additional spectral regions where such techniques need to be developed, but the pace has been quite rapid in the last few years. Maser oscillations in the infrared, optical, and ultraviolet regions have now been obtained in many ways and appear easy; new excitation mechanisms and systems are continually turning up. There are still two frequency regions, however, where such sources of radiation are rare or nonexistent. One is in the submillimeter region or far-infrared. The region has, in a sense, now been crossed and conquered by maser oscillators. But techniques in this spectral region are still rudimentary, and the frequency coverage with

some integer. If, then, the radiated frequencies are to have a fractional bandwidth of about $3 \cdot 10^{-17}$, such as would come from fundamental noise according to Eqn. 8, the mirror separation must not vary by more than this fractional amount. For a mirror separation of 1 meter, the motion allowed would be less than $3 \cdot 10^{-15}$ centimeter – a demanding requirement!

If the mirror separation is held constant by cylindrical rods, L must still vary as a result of thermal excitation of the lowest frequency-stretching modes of the rods. This gives an additional fluctuation which is usually larger than that from spontaneous emission. It produces a fractional motion[47]

$$\left(\frac{2kT}{YV}\right)^{\frac{1}{2}}$$

where T is the temperature, V the volume of the separators, and Y their Young's modulus.

In order to examine the monochromaticity of lasers, two helium–neon systems were carefully shock-mounted in an acoustically insulated wine cellar of an unoccupied and isolated house so that acoustic vibrations would be minimized[47]. Their pairs of mirrors were separated by heavy invar rods about 60 centimeters long. For this case, the limiting theoretical fluctuations set by thermal motions of the rods corresponded to fractional frequency variations of $5 \cdot 10^{-15}$, or a frequency fluctuation of 2 cycles per second. Light from each laser was sent into a photodetector, and the beat frequency examined electronically. Under good conditions free from acoustic disturbances or thermal transients in the invar spacers, this experiment showed that variations in the laser frequencies over periods of a few seconds were less than 20 cycles per second, or about one part in 10^{13}. This was ten times the limit of thermal fluctuations, but corresponded to detection of motions of the two mirrors as small as $5 \cdot 10^{-12}$ centimeter, a dimension comparable with nuclear diameters. Presumably, with great care, one can obtain results still nearer to the theoretical values.

The narrowest atomic spectral lines have widths of the order of 10^8 cycles per second, so that the laser measured was more monochromatic than earlier light sources by a factor of about 10^6. Light of this type can interfere with itself after traveling a distance of about 10000 kilometers. Hence it could in principle measure changes in such a large distance to a precision of one wavelength of light, if there were any optical path so constant. Interference work has been done in several laboratories with laser light over distances of a few hundred meters, which does not require quite such special elimination of acoustic or other disturbances.

A third property of laser light which is of interest is its directivity, or the spacial coherence across the beam. As indicated above, certain modes of oscillation should represent approximately a plane wave of cross section comparable with the mirror diameter D. The helium–neon maser seems to easily allow adjustment so that such a mode of oscillation occurs, and its beam has been shown[35,48] to have nearly the expected divergence λ/D due to diffraction.

The spacial coherence or planarity of a laser beam implies that the entire beam can be focused by a microscope to a region as small as about $\lambda/2$, or the resolving power of the microscope. Similarly, it may be transmitted through a telescope in a beam whose angular width is simply determined by the angular resolution of the telescope, and hence much less than the angular divergence λ/D as the beam emerges from a small laser. The entire energy is originally created in the ideal laser in a single mode; it can be transmitted into other single modes by optical systems without violating the well-known brightness laws of optics.

This brings us to a fourth important property, the intensity or brightness which can be achieved by maser techniques. As indicated initially, once one has the possibility of coherent amplification, there is no firm limit to intensity because equilibrium thermodynamics and Planck's law no longer are controlling. The only limit is set by the available energy input, heat dissipation, and size of the apparatus used.

If only the 1 milliwatt of power emitted by a helium–neon laser is focused by a good lens, the power density becomes high because the cross-sectional area of the focused spot would be only about $\lambda^2/4$. This gives a power density of $4 \cdot 10^5$ Watt/cm^2. The effective temperature of such a beam, because of its monochromaticity, is also rather high–about 10^{19} °K for the light of 20-cycle per-second bandwidth.

The pulsed systems, such as ruby lasers in particular, emit much greater power, although they do not quite approach the limits of coherence which the gaseous systems do. Ruby lasers emit a few tenths of a joule to a few hundred joules of energy in pulses from about 10^{-3} second to 10^{-8} second in length. The power can thus be as great as 10^9 Watts or more. Effective temperatures of the radiation are of the order 10^{23} °K. The actual limit of power density will generally be set by the limit of light intensity optical materials can stand without breakage or ionization. Power of 10^9 Watts focused to a spot 10^{-2} millimeter in diameter produces an electric field strength in the optical wave of about 10^9 Volt/cm, which is in the range of fields by which valence electrons are held in atoms. Hence this power ionizes and disrupts all material.

The radiation pressure also becomes large, being about 10^{12} dyne/cm^2, or 10^6 atmospheres, at such a focal point.

Some Applications of Lasers

It is clear that light in more ideal and in more intense form, which maser techniques have produced, can be expected to find application in wide and numerous areas of technology and of science simply because we find our present techniques of producing and controlling light already so widely applied. Most of these applications are still ahead of us, and there is not time to treat here even those which are already beginning to develop. I shall only mention that in technology lasers have been put to work in such diverse areas as radar, surgery, welding, surveying, and microscopy. A little more space will be devoted here to discussing three broad areas of science to which optical, infrared, and ultraviolet masers are expected to contribute.

Masers seem to provide the most precise techniques for measurement of the two fundamental dimensions of time and length. Over short periods of time maser oscillators clearly give the most oscillations; for longer times the hydrogen maser also seems to provide the most precise clock yet available. Light from optical masers allows new precision in the measurement of distance, and already seems capable of improving our standard of length. This new precision suggests interesting experiments on certain fundamental properties of our space, as well as the application of higher precision to a variety of physical effects. So far, experiments have been done to improve the precision with which the Lorentz transformation can be experimentally verified[49, 50]. It appears that improved precision in measurement of the speed of light can also be expected. If we look some distance in the future, it seems clear that the techniques of quantum electronics will allow direct measurement of the frequency of light, rather than only its wavelength. This can be accomplished by generation of harmonics of a radio frequency, amplification of the new frequency, and further generation of harmonics until the radio region is linked with optical frequencies. This should eventually allow measurement of the velocity of light, c, to whatever precision we define time and length. Or, it will allow the elimination of separate standards of time and of length because c times a standard time will define a standard length with more precision than we can now achieve.

The power of spectroscopy should be considerably increased by use of

masers. In particular, these very monochromatic sources can very much improve spectroscopic resolution and thus allow more detailed examination of the structure of atoms, molecules, or solids. This advance can be particularly striking in the infrared and far-infrared, where present resolution is far less than the widths of atomic or molecular lines. Already some high-resolution spectroscopy has been done with lasers[51,52] and still more interesting work of this general type can be expected before long.

A third interesting field for which lasers are important has emerged as a field almost entirely because of the existence of lasers, and is the area where scientific research has so far been most active. This is what is usually called nonlinear optics[53,54] although it includes some phenomena which might not previously have been described in this way. We have been accustomed in the past to discussing the progress of light through a passive optical material of more-ore-less fixed properties. But, in the intense laser beams now available, interactions between the light and the optical medium are sufficiently large that properties of the medium can no longer be regarded as fixed. The medium distorts, it molecules vibrate, and polarization of electrons in its atoms no longer responds linearly to the applied field. One must now also consider the dynamics medium, and interactions between their two motions. Some of the new phenomena observed are multiple-quanta absorption, which makes absorption depend on intensity[55,56], harmonic generation in optical materials and mixing of light frequencies[57–60], excitation of coherent molecular vibrations and stimulated Raman effects[42–45], and stimulated Brillouin scattering[61,62]. Only the last two of these will be discussed, partly because they bear on still another kind of maser, one which generates phonons.

The Phonon Maser

Acoustic waves follow equations that are of the same general form as the equations of light and manifest many of the same phenomena. An acoustic wave can produce an atomic or molecular excitation, or receive energy from it by either spontaneous or stimulated emission. Hence, one may expect maser action for acoustic waves if a system can be found in which molecules are sufficiently coupled to an acoustic field and appropriate excitation can be obtained to meet the threshold condition. The first such systems suggested involved inversion of the spin states of impurities in a crystal in ways similar to those used for solid-state electromagnetic masers[63]. A system of this type has

been shown to operate as expected[64]. However, a more generally applicable technique seems to be Brillouin scattering and its close associate Raman scattering, which utilize phase correlation rather than population inversion to produce amplification. This process can also be viewed as parametric amplification.

Light may be scattered by the train of crests and troughs in an acoustic wave much as by a grating. Since the wave is moving, the scattering involves a Doppler shift. The net result, first analyzed by Brillouin[65], is that the scattered light is shifted in frequency from the frequency ν_0 of the original beam by an amount

$$\nu = 2\nu_0 v/c \sin\theta/2 \qquad (13)$$

where v and c are the phase velocities of sound and of light, respectively, in the medium, and θ is the scattering angle. The energy lost, $h\nu$, is given to the scattering acoustic wave of frequency ν. If the light is of sufficient intensity, it can thus give energy to the acoustic field faster than it is lost and fulfill a threshold condition which allows the acoustic energy to build up steadily.

For the very high acoustic frequencies (10^9 to 10^{10} c/sec) implied by Eqn.13 when θ is not very small, the losses are usually so large that interesting amplification cannot be achieved with ordinary light. But, with laser beams of hundreds of megawatts per square centimeter, it is quite feasible to produce an intense build-up of acoustic waves by this process of stimulated Brillouin scattering[61,62] – so intense, in fact, that the acoustic energy can crack glass or quartz. This gives a method of producing and studying the behavior of very-high-frequency acoustic waves in almost any material which will transmit light – a possibility which was previously not so clearly available.

Brillouin scattering by spontaneous emission has been studied for some time. But the intense monochromatic light of lasers allows now much greater precision in work with this technique[52] and it too is yielding interesting information on the propagation of hypersonic waves in materials.

There is no firm limit to the acoustic frequencies which can be produced by stimulated emission, even though Eqn.13 indicates a kind of limit, for $\theta = \pi$, of $2\nu_0 v/c$. But in the optical branch of acoustic waves the phase velocity v can be very high. In fact, stimulated Raman scattering, or the Raman maser mentioned briefly above, represents excitation of the optical branches of acoustic spectra, and generates coherent molecular oscillations. Quantum-electronic techniques can thus allow interesting new ways to generate and explore most of the acoustic spectrum as well as much of the electromagnetic domain.

Concluding Remarks

In a few years this brief report will no longer be of much interest because it will be outdated and superseded, except for some matters of general principle or of historical interest. But, happily, it will be replaced by further striking progress and improved results. We can look forward to another decade of rapid development in the field of quantum electronics – new devices and unsuspected facets of the field, improved range and performance of masers, and extensive application to science and to technology. It seems about time now for masers and lasers to become everyday tools of science, and for the exploratory work which has demonstrated so many new possibilities to be increasingly replaced by much more finished, more systematic, and more penetrating applications. It is this stage of quantum electronics which should yield the real benefits made available by the new methods of dealing with radiation.

1. J.R. Pierce, *Phys. Today*, 3, No.11 (1950) 24.
2. N.F. Ramsey, *Phys. Rev.*, 103 (1956) 20.
3. W.E. Lamb and R.C. Retherford, *Phys. Rev.*, 79 (1950) 549.
4. E.M. Pyrcell and R.V. Pound, *Phys. Rev.*, 81 (1951) 279.
5. Bloch, Hansen and Packard, *Phys. Rev.*, 70 (1946) 474.
6. N.G. Basov and A.M. Prokhorov, *J. Exptl. Thoeret. Phys.* (*U.S.S.R.*), 27 (1954) 431.
7. J. Weber, *Inst. Electron. Elec. Engrs., Trans. Electron Devices*, 3 (1953) 1.
8. *Columbia Radiation Lab. Quart. Progr. Rept.*, (Dec.1951); C.H. Townes, *J. Inst. Elec. Commun. Eng.* (*Japan*), 36 (1953) 650 [in Japanese].
9. J.P. Gordon, H.J. Zeiger and C.H. Townes, *Phys. Rev.*, 95 (1954) 282.
10. J.P. Gordon, H.J. Zeiger and C.H. Townes, *Phys. Rev.*, 99 (1955) 1264.
11. R. Serber and C.H. Townes, in C.H. Townes (Ed.), *Quantum Electronics*, Columbia University Press, New York, 1960, p.233; see also H. Friedburg, *ibid.*, p.227.
12. Shimoda, Takahashi and Townes. J. *Phys. Soc.* (*Japan*), 12 (1957) 686; M.W. Muller, *Phys. Rev.*, 106 (1957) 8; M.W.P. Strandberg, *ibid.*, 106 (1957) 617; R.V. Pound, *Ann. Phys.* (*N.Y.*), 1 (1957) 24.
13. C.H. Townes, in J.R. Singer (Ed.), *Advances in Quantum Electronics*, Columbia University Press, New York, 1961, p.3.
14. Shimoda, Wang and Townes, *Phys. Rev.*, 102 (1956) 1308.
15. A.L. Schawlow and C.H. Townes, *Phys. Rev.*, 112 (1958) 1940.
16. A correction of a factor of 2 to δ given by Gordon, Zeiger and Townes[10], and Schawlow and Townes[15], is demonstrated by P. Grivet and A. Blaquiere, *Proc. Symp. Optical Masers*, Polytechnic Inst., Brooklyn, 1963.
17. Alsop, Giordmaine, Townes and Wang, *Phys. Rev.*, 107 (1957) 1450; J.P. Gordon and L.D. White, *ibid.*, 107 (1957) 1728.

18. J. de Prins and P. Kartaschoff, in *Proc. Course XVII, Enrico Fermi Intern. School of Physics*, Academic Press, New York, 1962, p. 88.
19. Goldenberg, Kleppner and Ramsey, *Phys. Rev. Letters*, 5 (1960) 361; Phys. Rev., 123 (1961) 530.
20. Combrisson, Honig and Townes, *Compt. Rend.*, 242 (1956) 2451; Feher, Gordon, Buehler, Gere and Thurmond, *Phys. Rev.*, 109 (1958) 221.
21. N. Bloembergen, *Phys. Rev.*, 104 (1956) 324.
22. N. G. Basov and A. M. Prokhorov, *J. Exptl. Theoret. Phys.* (*U. S. S. R.*), 28 (1955) 249.
23. Scovil, Feher and Seidel, *Phys. Rev.*, 105 (1957) 762.
24. Makhov, Kikuchi, Lambe and Terhune, *Phys. Rev.*, 109 (1958) 1399.
25. A. E. Siegman, *Microwave Solid State Masers*, McGraw-Hill, New York, 1964.
26. Alsop, Giordmaine, Mayer and Townes, *Astron. J.*, 63 (1958) 301; *Proc. IEEE* (*Inst. Electron. Elec. Engrs.*), 47 (1959) 1062.
27. E. Epstein, *Astron. J.*, 69 (1964) 490.
28. R. W. DeGrasse and H. E. D. Scovil, *J. Appl. Phys.*, 31 (1960) 443.
29. DeGrasse, Hogg, Ohm and Scovil, *J. Appl. Phys.*, 30 (1959) 2013.
30. A. L. Schawlow and C. H. Townes, *Phys. Rev.*, 112 (1958) 1940.
31. A. G. Fox and T. Li, *Bell System Tech. J.*, 40 (1961) 453.
32. Jacobs, Gould and Rabinowitz, *Phys. Rev. Letters*, 7 (1961) 415.
33. T. H. Maiman, *Nature*, 187 (1960) 493; Maiman, Hoskins, D'Haenens, Asawa and Evtuhov, *Phys. Rev.*, 123 (1961) 1151.
43. Collins, Nelson, Schawlow, Bond, Garrett and Kaiser, *Phys. Rev. Letters*, 5 (1960) 303.
35. Javan, Bennett and Herriott, *Phys. Rev. Letters*, 6 (1961) 106.
36. A. Javan, *Phys. Rev. Letters*, 3 (1959) 87.
37. R. J. Keyes and T. M. Quist, *Proc. IEEE* (*Inst. Electron. Elec. Engrs.*), 50 (1962) 1822.
38. Hall, Fenner, Kingsley, Soltys and Carlson, *Phys. Rev. Letters*, 9 (1962) 366.
39. Nathan, Dumke, Burns, Dill and Lasher, *Appl. Phys. Letters*, 1 (1962) 62.
40. Quist, Rediker, Keyes, Krag, Lax, McWhorter and Zeiger, *Appl. Phys. Letters*, 1 (1962) 91.
41. Basov, Bogdankevich and Devyatkov, *Dokl. Akad. Nauk S. S. S. R.*, 155 (1964) 783; *Proc. Symp. Radiative Recombination in Semiconductors, Paris*, 1964, Dunod Paris, 1965; C. Benoit à la Guillaume and T. M. Debever, *ibid.*
42. A. Javan, *J. Phys. Radium*, 19 (1958) 806; *Bull. Am. Phys. Soc.*, 3 (1958) 213; *Proc. Course XXXI, Enrico Fermi Intern. School of Physics*, Academic Press, New York, 1964.
43. E. J. Woodbury and W. K. Ng, *Proc. IEEE* (*Inst. Electron. Elec. Engrs.*), 50 (1962) 2367; Eckhardt, Hellwarth, McClung, Schwarz, Weiner and Woodbury, *Phys. Rev. Letters*, 9 (1962) 455.
44. R. W. Terhune, *Solid State Design*, 4 (1963) 38; Minck, Terhune and Rado, *Appl. Phys. Letters*, 3 (1963) 181; see also B. Stoicheff, *Proc. Course XXXI, Enrico Fermi Intern. School of Physics*, Academic Press, New York, 1964.
45. R. W. Hellwarth, *Phys. Rev.*, 130 (1963) 1850; H. J. Zeiger and P. E. Tannenwald, in P. Grivet and N. Bloembergen (Eds.), *Quantum Electronics: Proc. 3rd Intern. Congr.*, Columbia University Press, New York, 1964, p. 1589; Garmire, Pandarese and Townes, *Phys. Rev. Letters*, 11 (1963) 160.

46. E.L. Tolnas, *Bull. Am. Phys. Soc.*, 5 (1960) 342.
47. Jaseja, Javan and Townes, *Phys. Letters*, 10 (1963) 165.
48. D. Herriott, *J. Opt. Soc. Am.*, 52 (1962) 31.
49. J.P. Cedarholm and C.H. Townes, *Nature*, 184 (1959) 1350.
50. Jaseja, Javan, Murray and Townes, *Phys. Rev.*, 133 (1964) A1221.
51. A. Szöke and A. Javan, *Phys. Rev. Letters*, 10 (1963) 521.
52. R. Chiao and B. Stoicheff, *J. Opt. Soc. Am.*, 54 (1964) 1286; Benedek, Lastovka, Fritsch and Greytak, *ibid.*, 54 (1964) 1284; Masch, Starunov and Fabelinskii, *J. Exptl. Theoret. Phys. (U. S. S. R.)*, 47 (1964) 783.
53. C. A. Akhmanov and R. B. Khoklov, *Problems in Nonlinear Optics, 1962–1963*, Academy of Sciences, U.S.S.R., Moscow, 1964.
54. N. Bloembergen, *Nonlinear Optics*, Benjamin, New York, 1965.
55. I. Abella, *Phys. Rev. Letters*, 9 (1962) 453.
56. J. A. Giordmaine and J. A. Howe, *Phys. Rev. Letters*, 11 (1963) 207.
57. Franken, Hill, Peters and Weinreich, *Phys. Rev. Letters*, 7 (1961) 118; P. A. Franken and J. F. Ward, *Rev. Mod. Phys.*, 35 (1963) 23.
58. J. A. Giordmaine, *Phys. Rev. Letters*, 8 (1962) 19; Terhune, Maker and Savage, *Appl. Phys. Letters*, 2 (1963) 54.
59. J. Ducuing and N. Bloembergen, *Phys. Rev. Letters*, 10 (1963) 474.
60. Armstrong, Bloembergen, Ducuing and Pershan, *Phys. Rev.*, 127 (1962) 1918; N. Bloembergen and P. S. Pershan, *ibid.*, 128 (1962) 606.
61. Chiao, Garmire and Townes, *Proc. Course XXXI, Enrico Fermi Intern. School of Physics*, Academic Press, New York, 1964.
62. Chiao, Townes and Stoicheff, *Phys. Rev. Letters*, 12 (1964) 592.
63. C.H. Townes, in C.H. Townes (Ed.), *Quantum Electronics*, Columbia University Press, New York, 1960, p. 402; C. Kittel, *Phys. Rev. Letters*, 6 (1961) 449.
64. E.B. Tucker, *Phys. Rev. Letters*, 6 (1961) 547.
65. L. Brillouin, *Ann. Phys. (Paris)*, 17 (1922) 88; M. Born and K. Wang, *Dynamical Theory of Crystal Lattices*, Oxford University Press, New York, 1964.

Biography

Charles Hard Townes was born in Greenville, South Carolina, on July 28, 1915, the son of Henry Keith Townes, an attorney, and Ellen (Hard) Townes. He attended the Greenville public schools and then Furman University in Greenville, where he completed the requirements for the Bachelor of Science degree in physics and the Bachelor of Arts degree in Modern Languages, graduating *summa cum laude* in 1935, at the age of 19. Physics had fascinated him since his first course in the subject during his sophomore year in college because of its «beautifully logical structure». He was also interested in natural history while at Furman, serving as curator of the museum, and working during the summers as collector for Furman's biology camp. In addition, he was busy with other activities, including the swimming team, the college newspaper and the football band.

Townes completed work for the Master of Arts degree in Physics at Duke University in 1936, and then entered graduate school at the California Institute of Technology, where he received the Ph. D. degree in 1939 with a thesis on isotope separation and nuclear spins.

A member of the technical staff of Bell Telephone Laboratories from 1933 to 1947, Dr. Townes worked extensively during World War II in designing radar bombing systems and has a number of patents in related technology. From this he turned his attention to applying the microwave technique of wartime radar research to spectroscopy, which he foresaw as providing a powerful new tool for the study of the structure of atoms and molecules and as a potential new basis for controlling electromagnetic waves.

At Columbia University, where he was appointed to the faculty in 1948, he continued research in microwave physics, particularly studying the interactions between microwaves and molecules, and using microwave spectra for the study of the structure of molecules, atoms, and nuclei. In 1951, Dr. Townes conceived the idea of the maser, and a few months later he and his associates began working on a device using ammonia gas as the active medium. In early 1954, the first amplification and generation of electromagnetic waves by stimulated emission were obtained. Dr. Townes and his students coined

the word «maser» for this device, which is an acronym for *m*icrowave *a*mplification by *s*timulated *e*mission of *r*adiation. In 1958, Dr. Townes and his brother-in-law, Dr. A. L. Schawlow, now of Stanford University, showed theoretically that masers could be made to operate in the optical and infrared region and proposed how this could be accomplished in particular systems. This work resulted in their joint paper on optical and infrared masers, or lasers (*l*ight *a*mplification by *s*timulated *e*mission of *r*adiation). Other research has been in the fields of radio astronomy and nonlinear optics.

Having joined the faculty at Columbia University as Associate Professor of Physics in 1948, Townes was appointed Professor in 1950. He served as Executive Director of the Columbia Radiation Laboratory from 1950 to 1952 and was Chairman of the Physics Department from 1952 to 1955.

From 1959 to 1961, he was on leave of absence from Columbia University to serve as Vice President and Director of Research of the Institute for Defense Analyses in Washington, D. C., a nonprofit organization operated by eleven universities.

In 1961, Dr. Townes was appointed Provost and Professor of Physics at the Massachusetts Institute of Technology. As Provost he shared with the President responsibility for general supervision of the educational and research programs of the Institute. In 1966, he became Institute Professor at M. I. T., and later in the same year resigned from the position of Provost in order to return to more intensive research, particularly in the fields of quantum electronics and astronomy. He was appointed University Professor at the University of California in 1967. In this position Dr. Townes is participating in teaching, research, and other activities on several campuses of the University, although he is located at the Berkeley campus.

During 1955 and 1956, Townes was a Guggenheim Fellow and a Fulbright Lecturer, first at the University of Paris and then at the University of Tokyo. He was National Lecturer for Sigma Xi and also taught during summer sessions at the University of Michigan and at the Enrico Fermi International School of Physics in Italy, serving as Director for a session in 1963 on coherent light. In the fall of 1963, he was Scott Lecture at the University of Toronto.

Dr. Townes has served on a number of scientific committees advising governmental agencies and has been active in professional societies. He and his wife (the former Frances H. Brown; they married in 1941) live at 1988 San Antonio Avenue, Berkeley, California. They have four daughters, Linda Rosenwein, Ellen Anderson, Carla Lumsden, and Holly.

NIKOLAI G. BASOV

Semiconductor lasers

Nobel Lecture, December 11, 1964

In modern physics, and perhaps this was true earlier, there are two different trends. One group of physicists has the aim of investigating new regularities and solving existing contradictions. They believe the result of their work to be a theory; in particular, the creation of the mathematical apparatus of modern physics. As a by-product there appear new principles for constructing devices, physical devices.

The other group, on the contrary, seeks to create physical devices using new physical principles. They try to avoid the inevitable difficulties and contradictions on the way to achieving that purpose. This group considers various hypotheses and theories to be the by-product of their activity.

Both groups have made outstanding achievements. Each group creates a nutrient medium for the other and therefore they are unable to exist without one another; although, their attitude towards each other is often rather critical. The first group calls the second «inventors», while the second group accuses the first of abstractness or sometimes of aimlessness. One may think at first sight that we are speaking about experimentors and theoreticians. However, this is not so, because both groups include these two kinds of physicists.

At present this division into two groups has become so pronounced that one may easily attribute whole branches of science to the first or to the second group, although there are some fields of physics where both groups work together.

Included in the first group are most research workers in such fields as quantum electrodynamics, the theory of elementary particles, many branches of nuclear physics, gravitation, cosmology, and solid-state physics.

Striking examples of the second group are physicists engaged in thermonuclear research, and in the fields of quantum and semiconductor electronics.

Despite the fact that the second group of physicists strives to create a physical device, their work is usually characterized by preliminary theoretical analysis. Thus, in quantum electronics, there was predicted theoretically the possibility of creating quantum oscillators: in general, also, there were predicted the high monochromaticity and stability of the frequency of masers,

the high sensitivity of quantum amplifiers, and there was investigated the possibility of the creation of various types of lasers.

This lecture is devoted to the youngest branch of quantum electronics – semiconductor lasers, which was created only two years ago, although a theoretical analysis started already in 1957 preceded the creation of lasers[1].

However, before starting to discuss the principles of operation of semiconductor lasers we would like to make some remarks of the theoretical «by-products» of quantum electronics. There are many of them but we shall consider only three:

(*1*) The creation of quantum frequency oscillators of high stability and the transition to atomic standards of time made it possible to raise the question of solving the problem of the properties of atomic time.

Dicke[2] in his paper at the first conference on quantum electronics pointed out the possibility of an experimental check of the hypothesis on the variation of fundamental physical constants with time on the basis of studying changes in frequencies of different quantum standards with time. There arises the question about the maximum accuracy of atomic and molecular clocks depending on the nature of quantum of emission, especially about the accuracy of the measurement of short time intervals.

(*2*) Due to quantum electronics there was started an intensive investigation of a new «super non-equilibrium state of matter» – the state with negative temperature, which in its extreme state of negative zero is close in its properties to the absolute ordering intrinsic for the temperature of absolute zero. It is just this property of high ordering of a system with negative temperature which makes it possible to produce high-coherent emission in quantum oscillators, to produce high sensitive quantum amplifiers, and to separate the energy stored in the state with negative temperature in a very short time, of the order of the reciprocal of the emission frequency.

(*3*) Quantum electronics gives examples of systems in which there occurs radiation with a very small value of entropy. For instance, spontaneous low temperature radiation from flash tubes, distributed through very large number of degrees of freedom is converted with the help of a system in a state of negative temperature (quantum oscillators) into high-coherent laser emission, the temperature of which in present experiments already attains a value of 10^{20} degrees.

Apparently, the regularities established by quantum electronics for radiation may be generalized for other natural phenomena. The possibility of obtaining a high degree of organization with the help of feed-back systems may

be of interest for chemical and biological research, and for cosmology. The question arises as to whether or not the maser principle is used in Nature.

We believe that the above questions need attention from physicists of the first group, because these questions go beyond the limits of the theory of oscillations, the theory of radiation and usual optics which form the basis of modern quantum electronics.

I. Conditions for the Production of Negative Temperature in Semiconductors

Investigations of semiconductor quantum oscillators were a direct continuation of research on molecular oscillators and paramagnetic amplifiers. One should note that at the beginning of research on semiconductor lasers, due to investigations in the field of semiconductor electronics, there became known the physical characteristics of semiconductors, which were essential in the development and practical realization of lasers; such as, optical and electric characteristics, structure of energy bands, and relaxation time.

Various pure and alloyed semiconductors were made, and the technique of measurements of their various properties and the technology of making p–n junctions were worked out. All of this considerably simplified investigations of semiconductor lasers. Semiconductors were very intriguing because of the possibility of using them for making oscillators with a frequency range from the far-infrared region to the optical or even to the ultraviolet range, as well as because of the variety of methods by means of which states with negative temperatures may be obtained within them and because of their large factor of absorption (amplification). As the following studies have shown, semiconductor lasers may have extremely high efficiency, in some cases approximating 100 percent.

In contrast to an isolated atom, in semiconductors there do not exist separate energy levels, but rather there exist groups of energy levels arranged very close to one another, which are called bands (Fig. 1). The upper group of levels, called the conduction band, and a lower group of excited levels, called the valent band, are divided by a band of forbidden energy (Fig. 1).

The distribution of electrons on energy levels is described by the Fermi function: each level is occupied by two electrons, the electrons being distributed in the energy range of the order of the energy of kT thermal motion; and, the probability of finding an electron beyond the kT interval sharply decreases when the energy level increases. If the energy of thermal motion is

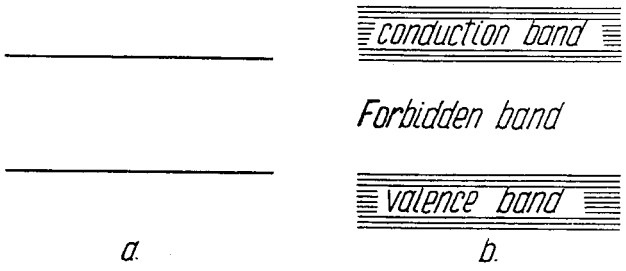


Fig. 1. Energy-level diagram. (a) For atoms with two energy levels; (b) for semiconductors.

significantly less than the energy difference between the conduction and valent bands, then practically all electrons will be found in the valent band, filling its levels, while practically all levels of the conduction band remain free (Fig. 2a). In such a state the semiconductor cannot conduct electric current and becomes an insulator, since the electric field applied to the semiconductor is unable to change the motion of the electrons in the valent band (all energy levels are occupied).

If the energy of thermal motion is sufficient, then a part of the electrons are thrown into the conduction band. Such a system may serve as a conductor of electric current. Current is able to flow both due to variation of the electron energy under the action of the external field, as well as due to changes in the electron distribution within the valent and conduction bands. Current within the valent band behaves as if those places free from electrons (holes) moved in a direction opposite to that of the electrons. A vacant place or «hole» is entirely equivalent to a positively charged particle (Fig. 2b).

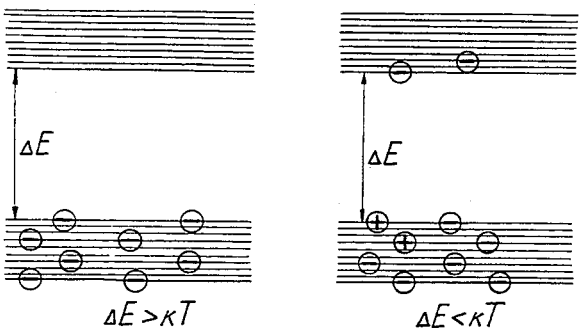


Fig. 2. Distribution of the electrons on energy levels.

During interaction with light, a semiconductor, similar to an isolated atom, may undergo three processes:

(1) A light quantum may be absorbed by the semiconductor: and, in this case an electron–hole pair is produced. The difference in energy between the electron and the hole is equal to the quantum energy. This process is connected with the decrease in energy of the electromagnetic field and is called *resonance absorption* (Fig. 3a).

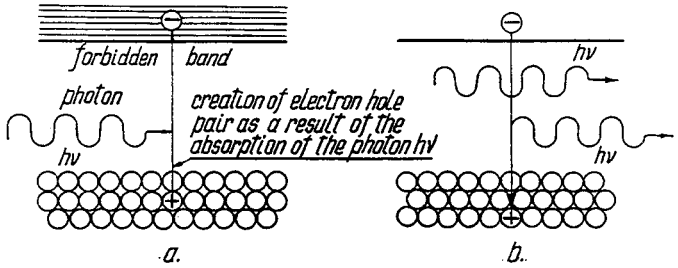


Fig. 3. Processes of the interaction with light. (a) Resonance absorption; (b) stimulated emission.

(2) Under the influence of a quantum, an electron may be transferred from the conduction band to a vacant place (hole) on the valent band. Such a transfer will be accompanied by the emission of a light quantum identical in frequency, direction of propagation and polarization to the quantum which produced the emission. This process is connected with an increase of the field energy and is called *stimulated emission* (Fig. 3b). We recall that stimulated emission was discovered by A. Einstein in 1917 during an investigation of thermodynamical equilibrium between the radiation field and atoms.

(3) Besides resonance absorption and stimulated emission, a third process may take place–*spontaneous emission*. An electron may move over to a vacant place–hole (recombine with the hole) in the absence of any radiation quanta.

Since the probabilities of stimulated radiation and resonance absorption are exactly equal to one another, a semiconductor in an equilibrium state at any temperature may only absorb light quanta, because the probability of finding electrons at high levels decreases as the energy increases. In order to make the semiconductor amplify electromagnetic radiation, one must disturb the equilibrium of the distribution of electrons within the levels and artificially produce a distribution where the probability of finding electrons on higher energy

levels is greater than that of finding them on the lower levels[1, 3]. It is very difficult to disturb the distribution inside a band because of the strong interaction between the electrons and the lattice of the semiconductor: it is restored in 10^{-10} to 10^{-12} sec. It is much simpler to disturb the equilibrium between the bands, since the lifetime of electrons and holes is considerably greater in the bands. It depends on the semiconductor material and lies in the interval of 10^{-3} to 10^{-9} sec.

Due to the fact that electrons and holes move in semiconductors, in addition to the law of the conservation of energy, the law of the conservation of momentum should be fulfilled during emission. Since the photon impulse is extremely small, the law of the conservation of momentum, approximately speaking, requires that the electrons and holes must have the same velocity during the emission (or absorption) of a light quantum. Fig. 4 shows graphi-

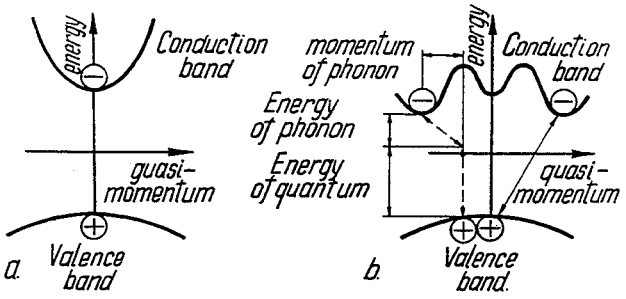


Fig. 4. Diagram of the electron–hole energy dependence on the quasi-momentum. (a) Direct transitions; (b) indirect transitions.

cally the dependence of energy on momentum. There are two types of semiconductors. For one group of semiconductors, the minimum of electron energy in the conduction band is exactly equal to the maximum of hole energy in the valent band. In such semiconductors there may take place so called «direct transitions». An electron having minimum energy may recombine with a hole having maximum energy. For another group of semiconductors, the minimum energy in the conduction band does not coincide with the maximum energy in the valent band. In this case the process of emission or absorption of a light quantum should be accompanied by a change in the amplitude of the oscillatory state of the crystal lattice, that is by the emission or absorption of a phonon which should compensate for the change in momentum. Such processes are called indirect transitions. The probability of indirect transitions is usually less than that of direct transitions.

In order to make a semiconductor amplify incident radiation under inter-band transitions, one should distinguish two cases:

(*a*) *In the case of direct transitions*

It is necessary to fill more than half of the levels in the band of the order of kT near the band's edge with electrons and holes. Such states, both for atoms and molecules, came to be called states with inverse populations, or states with negative temperature. The distribution of electrons when all levels in the kT zone of the conduction band are occupied by electrons, and in the valent band – by holes, corresponds to the temperature minus zero degrees. In this state (in contrast to the state of plus zero degrees), the semiconductor is only able to emit (stimulated and spontaneous) light quanta and is unable to absorb emissions.

The state of a semiconductor when most levels in a certain energy band are occupied by electrons or holes was named the *degenerated state*.

Thus, for the creation of negative temperature there must occur degeneration of electrons and holes in the semiconductor. With a given number of electrons and holes it is always possible to produce degeneration by means of lowering the semiconductor's temperature; since, as the temperature decreases the energy band width occupied by the electrons also decreases. At the temperature of liquid nitrogen for degeneration to take place it is necessary to have an electron concentration[3] of 10^{17}–10^{18} l/cm^3.

(*b*) *In the case of indirect transitions*

Degeneration is not necessary for the creation of negative temperature. This is connected with the fact that when indirect transitions occur, the probability of quantum-stimulated emission may not be equal to the probability of resonance absorption.

Consider, for instance, an indirect transition in which a quantum and a phonon are emitted simultaneously. The process of the simultaneous absorption of a quantum and a phonon is the inverse of that process.

The probability of absorption is proportional to the number of phonons in the crystal lattice. The number of phonons decreases with a lowering of temperature. At low temperature phonons are absent. Therefore, by means of lowering the temperature of the sample one may make the probability of emission much greater than the probability of absorption. This means that

with indirect transitions negative temperature may be attained with a considerably lower concentration of electrons and holes[4].

One should note that the absorption and emission of quanta during transitions within a band also takes place due to indirect transitions. When negative temperature is created between bands, the distribution of electrons (and holes) within a band corresponds to a positive temperature and leads to the absorption of emission.

In the case of direct transitions, when the probability of interband transitions is much greater than that of innerband transitions, one may neglect the innerband transitions; that is, states with negative temperature can be used for the amplification of emission.

In the case of indirect transitions for amplification to take place it is not sufficient to attain negative temperature. It is necessary that the probability of interband transitions be greater than that of innerband transitions. The necessity of fulfilling this condition makes it difficult to utilize indirect transitions. According to Dumke's estimate, this condition cannot be fulfilled for germanium[5]. However, it may be fulfilled for other semiconductors[6].

In a number of cases in semiconductors, an electron and a hole form an interconnected state something like an atom-exciton. The excitons may recombine, producing an emission. They may be also used to obtain quantum amplifiers, but we shall not consider this in detail.

We have studied conditions for the production of negative temperature in semiconductors possessing an ideal lattice. In a non-ideal crystal there occur additional energy levels connected with various disturbances in the crystalline lattice (impurities, vacancies, dislocations, etc.). As a rule, these states are localized near the corresponding centre (for instance, near an impurity atom) and in this they differ from those states in the valent and conduction bands which belong to the crystal as a whole.

In an ideal crystal the number of electrons in the conduction band is exactly equal to the number of holes in the valent band. However, in an actual crystal the number of current carriers – electrons and holes – is determined, mainly, by the existence of impurities (Fig. 5).

There are two kinds of impurities: one type has energy levels arranged near the conduction band and creates excess electrons due to thermal ionization. These are called «donor» impurities. Other impurities having energy levels near the valent band are able of removing electrons from the valent band and thus producing an excess number of holes in it. These impurities are called «acceptors».

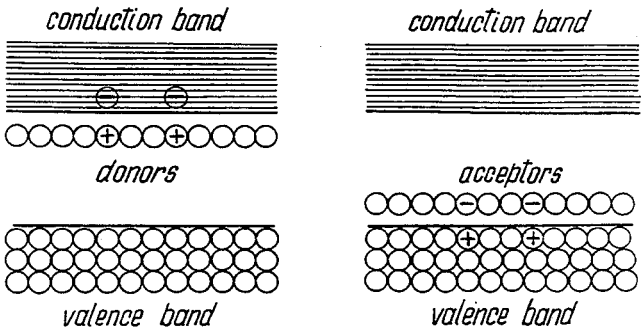


Fig. 5. Donor and acceptor levels.

One should note that a semiconductor with an equal number of donor and acceptor impurities behaves as if it were a pure semiconductor, since the holes produced by acceptors recombine with the electrons produced by donors.

In a number of cases, transitions of electrons between bands, between impurity atoms or between other levels may also be accompanied by emission of photons. One may likewise use these transitions for the creation of negative temperature. However, because of time limitations, we shall not discuss this question.

II. Methods of Obtaining States with Negative Temperature in Semiconductors

(*a*) *The method of optical pumping*

In the case of semiconductors one may utilize the «three-level» scheme[7] which has been used successfully for paramagnetic quantum amplifiers[8] and optical generators based on luminescent crystals and glasses[9] (Fig. 6).

Since the relaxation time of electrons and holes in the band levels[10] is much less than the lifetime of electrons and holes in the corresponding bands, one may obtain an inverse population by means of optical pumping.

Semiconductors have a very large absorption index which sharply increases as the radiation frequency increases. Therefore, to obtain an inverse population in samples of relatively large thickness, it is reasonable to use monochromatic radiation with a frequency close to that of the interband transitions[11]. In the case when the frequency of the exciting radiation is greater than the width of the forbidden band, a state with negative temperature is produced in a narrow band, several microns deep (on the order of the electrons' diffusion length)

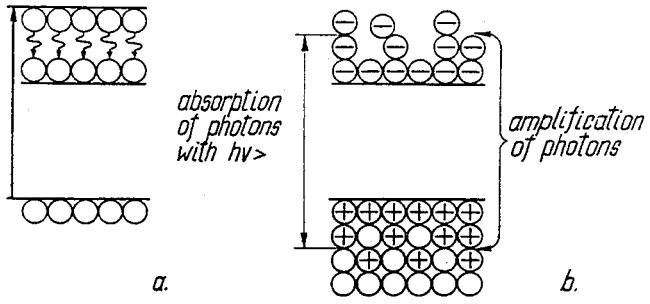


Fig. 6. Optical pumping. (a) Three levels diagram for atoms; (b) for semiconductors.

near the surface of the sample. As a source of radiation one may use the light from other types of lasers: gas lasers, lasers based on luminescent crystals or lasers based on p–n junctions[11].

(*b*) *The excitation of semiconductors by a beam of fast electrons*

If a beam of fast electrons is directed into the surface of a semiconductor, the electrons easily penetrate into the semiconductor. On their way the electrons collide with the atoms of the crystal and create electron–hole pairs. Calculations and experiments[12,13] have shown that an amount of energy approximately three times greater than the minimum energy difference between the bands is spent on the production of one electron–hole pair (Fig. 7a). The electrons and holes obtained give their excess energy to the atoms of the lattice and accumulate in the levels near the edges of the corresponding bands. In this case a state with negative temperature may be created[14,15]. The higher the electron energy, the deeper they will penetrate. However, there exists a

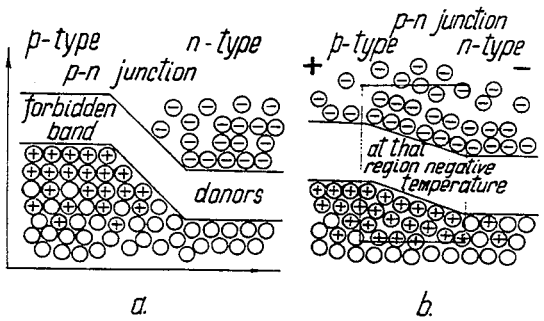


Fig. 7. (a) p–n junctional equilibrium; (b) p–n junction in the external electrical field.

certain threshold energy, beginning with which the electrons will produce defects in the crystal; that is, will destroy the crystalline lattice. This threshold energy depends upon the binding energy of the atoms in the crystals and is usually equal to about several hundred KeV. Experiments have shown that electrons with energy in the range of 200–500 KeV are not yet capable of noticeably harming the lattice.

The current density of fast electrons at which negative temperature is produced strongly depends upon the lifetime of electrons and holes. For semiconductors with a lifetime of 10^{-7} sec at the temperature of liquid nitrogen, the threshold of the current density has the order of one Ampere per cm^2. Since in the presence of such large currents it is difficult to remove the energy released in the semiconductor, the impulse method of excitation with a short impulse duration is usually employed.

(*c*) *The injection of electrons and holes through p–n junctions*

As it was noted above, a specific characteristic of semiconductors is that their energy levels may be filled with electrons or holes by introducing into the crystals special types of impurity atoms. However, the simultaneous introduction of donor and acceptor impurities does not result in the production of states with negative temperature. Therefore, in order to obtain an inverse population one does as follows: take two pieces of a semiconductor, inject donor impurities into one of them, and inject acceptor impurities into the other. If one then connects one piece to the other, a *p–n* junction will be created. On the boundary between the semiconductors there arises a potential difference which does not allow electrons to penetrate into the crystal having holes and likewise does not allow holes to penetrate into the crystal having electrons (Fig. 7a). As it was pointed out above, a large concentration of electrons and holes is necessary for the production of an inverse population (more than half of the levels in a certain energy band should be occupied), that is the semiconductor must have a large number of impurities.

If one applies an external voltage to a *p–n* junction, removing the potential difference between the two pieces of the semiconductor, the equilibrium of the distribution of electrons will be disturbed, and current will flow through the semiconductor. In this case electrons appear to flow into the region with a large concentration of holes, and holes – into the region with a large concentration of electrons. An inverse population arises in a narrow region near the *p–n* junction at a distance of several microns. Thus, there is obtained a layer of

the semiconductor which is able to amplify electromagnetic waves by means of the stimulated emission of quanta during the transition of electrons from the conduction band to the valent band[16] (Fig. 7b).

Many methods for the production of p–n junctions were worked out during research on semiconductors. At the present time two methods of making p–n junctions are used for the creation of lasers: the diffusion method[17,18] and the method of dopping with different impurities during the process of growing a crystal[19].

III. Semiconductor Lasers

In order to carry out generation on the basis of systems with negative temperature, one must introduce feedback coupling into the system. This feedback coupling is carried out with the aid of cavities. The simpliest type of cavity in the optical range is a cavity with plane-parallel mirrors[20,21]. Light quanta reflecting from the mirrors will pass many times through the amplifying medium. If a light quantum, before its absorption by the mirrors or inside the sample, has time to cause stimulated emission of more than one quantum (that is, if the condition of self-excitation is fulfilled in the system), that system will operate as a laser (Fig. 8). If one maintains a certain negative

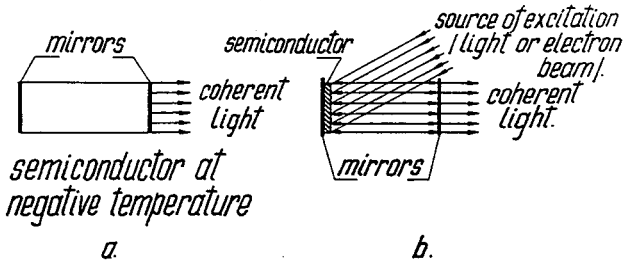


Fig. 8. Diagram of semiconductor lasers. (a) Usual; (b) with radiative mirrors.

temperature in the sample with the help of an external energy source, the number of quanta in the cavity will increase until the quantity of electrons excited per time unit becomes equal to the number of emitted quanta.

It should be especially noted, that when a quantum system with feed-back coupling operates as a laser, its emission has a very narrow frequency band. This characteristic makes laser emission different from all other light sources:

filament lamps, luminescent lamps and light sources with very narrow atomic and molecular spectral lines.

The monochromatic emission of lasers is a result of the properties of stimulated radiation: the quantum frequency of the stimulated radiation equals the frequency of the quantum which produced the radiation. The initial line width in semiconductors is usually about several hundred ångstroms. At the present time it has been shown that the line width in lasers which use a *p–n* junction in GaAs is less than fifty megacycles[22, 24]. The minimum value of the line width in lasers is connected with the phenomenon of spontaneous emission.

Spatial directivity of the emission arises together with change in the spectral composition of the oscillation regime. It is connected also with the nature of stimulated radiation: during stimulated radiation, a light quantum has the same direction of propagation as the quantum which produced it.

Usually in semiconductor lasers, the sample itself serves as a cavity; since semiconductor crystals have a large dielectric constant, and, since the polished boundary of the division between the air and the dielectric is able of reflecting about 30% of the radiation.

The first semiconductor lasers were made utilizing *p–n* junctions in crystals of GaAs[17,18]. Some time later, lasers were made under excitation by an electron beam[15], and, recently, under excitation by a light beam[23]. In Table 1 different semiconducting materials are shown with which lasers have been made, and the methods of excitation are given.

With the help of semiconductors it has already become possible to cover a large frequency range from 0.5μ to 8.5μ. In a number of cases it is possible to continuously overlap a very large frequency range, since the variation of the concentration of components in three-component semiconductors units results in changes in the distances between the bands, that is, allows one to continuously change the emission frequency. For instance, variation of composition in the system InAs–InP results[25] in frequency changes from 0.9μ to 3.2μ.

At present, the highest degree of development has been obtained with lasers utilizing *p–n* junctions in GaAs. Impulse and continuous regimes were obtained with an average power of several Watts, and peak power of up to 100 Watts, with an efficiency[24] of about 30%.

The most interesting characteristic of semiconductor lasers is their high efficiency.

Since a direct transformation of electric current into coherent emission takes place in lasers utilizing *p–n* junctions, their efficiency may approach uni-

Table 1
Semiconductor lasers

The semiconductor material	*The wave range of the radiation (in microns)*	*The method of the excitation*	*References*
CdS	0.5	high speed electron beam	15
CdTe	0.8	high speed electron beam	30
GaAs	0.85	*p–n* junction	17, 18
		high speed electron beam	29
		optical excitation	23
InP	0.9	*p–n* junction	31
GaSb	1.6	*p–n* junction	32
		high speed electron beam	33
InAs	3.2	*p–n* junction	34
		high speed electron beam	35
InSb	5.3	*p–n* junction	36
		high speed electron beam	6
PbTe	6.5	*p–n* junction	24
PbSe	8.5	*p–n* junction	24
GaAs–GaP	0.65–0.9	*p–n* junction	37
InAs–InP	0.9–3.2	*p–n* junction	25
GaAs–InAs	0.85–3.2	*p–n* junction	38

ty. Even now it has become possible to make diodes with an efficiency [26] of 70–80%.

Lasers with monochromatic optical pumping should also have a very high efficiency, since the pumping frequency may be made close to the emission frequency[11].

The efficiency of lasers with electron excitation cannot be higher[12] than about 30%, since two thirds of the energy is spent on heating of the lattice during the production of electron–hole pairs. However, such lasers may be rather powerful. This type of excitation will evidently make it possible to create sources of coherent emission working in the far ultraviolet range.

Another characteristic of semiconductors is a high coefficient of amplification, attaining a value of several thousands of reverse centimeters, which makes possible to construct lasers with dimensions measured in microns, that is, with cavity dimensions close to the length of the emission wave. Such cavities should have a very short setup time, of the order of 10^{-12}–10^{-13} sec, which opens the way for the control of high frequencies by using the oscillations in semiconductor lasers, and for the creation of superfast-operating circuits on

the basis of lasers, such as components for superfast-operating electronic computers. Q-switch lasers giving very short light pulses may be built out of semiconducting materials.

The small dimensions of semiconductor lasers make it possible to construct quantum amplifiers with an extremely high sensitivity, since sensitivity increases with a decrease in the number of modes of oscillation which may be excited in the cavity. For the first time light amplifiers with an amplification index of about 2000 cm^{-1} have been produced[28].

The high amplification index in semiconductor lasers makes it possible to create for them a new type of cavity – the cavity with emitting mirrors (Fig. 8)[27].

A silver mirror is covered by a thin semiconductor film which is then covered by a transparent film. If one produces in the semiconducting film a state with negative temperature which can compensate for the mirror losses, such a mirror may be used in the construction of a laser. As in the case of a gas laser, one may expect to observe very high monochromaticity and spatial coherence in the emission. A significant advantage of such a system is the simplicity of removing heat from the thin semiconducting film, which indicates that it should be possible to obtain considerable power.

In order to produce negative temperature in a semiconducting film, one may use electronic excitation or optical pumping. The utilization of semiconductor lasers with *p–n* junctions for optical pumping makes it possible to attain high efficiency in the system as a whole.

The question as to the maximum power which may be obtained using semiconductor lasers is not quite clear at present. However, the employment of emitting mirrors of sufficiently large area will make it possible, apparently, to utilize a considerable quantity of semiconducting material. The maximum value of a mirror's cross-section is determined by such factors as the precision of its manufacture and the homogeneity of its semiconducting layer. Various deviations from optical homogeneity will produce the highest modes of oscillations.

Among the disadvantages of semiconductor lasers are their relatively small power, their large spatial divergence and their insufficiently high monochromaticity.

However, in speaking about those disadvantages one should keep in mind that the field of semiconductor quantum electronics is still in its infancy. Furthermore, the means of overcoming these disadvantages are already in sight. It is quite clear in what directions to proceed in order to develop semi-

conductor quantum electronics, and to increase the sphere of application of semiconductor lasers. All of this gives reason to hope that semiconductor quantum electronics will continue to play a fundamental role in the development of lasers.

1. N.G.Basov, B.M.Vul and Yu.M.Popov, *Soviet JETP (U.S.S.R.)*, 37 (1959) 585.
2. R.H.Dicke, *Quantum Electronics*, Columbia University Press, New York, 1960, p.572.
3. N.G.Basov, O.N.Krokhin and Yu.M.Popov, *Usp.Fiz.Nauk*, 72 (1960) 161.
4. N.G.Basov, O.N.Krokhin and Yu.M.Popov, *Soviet JETP (U.S.S.R.)*, 39 (1960) 1001.
5. W.P.Dumke, *Phys.Rev.*, 127 (1962) 1559.
6. C.Benoit à la Guillaume and J.M.Debever, *Proc.Symp.Radiative Recombination in Semiconductors*, Paris, 1964, Dunod, Paris, 1965.
7. N.G.Basov and A.M.Prochorov, *Soviet JETP (U.S.S.R.)*, 28 (1955) 249.
8. N.Bloembergen, *Phys.Rev.*, 104 (1956) 324.
9. T.H.Maiman, *Nature*, 187 (1960) 493.
10. O.N.Krokhin and Yu.M.Popov, *Soviet JETP (U.S.S.R.)*, 38 (1960) 1589.
11. N.G.Basov and O.N.Krokhin, *Soviet JETP (U.S.S.R.)*, 46 (1964) 1508.
12. Yu.M.Popov, *Proc.FIAN*, 23 (1963) 67.
13. V.S.Vavilov, *Usp.Fiz.Nauk*, 25 (1961) 263.
14. N.G.Basov, O.N.Krokhin and Yu.M.Popov, *Advan.Quant.Elec.*, (1961) 496.
15. N.G.Basov, O.V.Bogdankevich and A.Devyatkov, *Dokl.Akad.Nauk (S.S.S.R.)*, 155 (1964) 783.
16. N.G.Basov, O.N.Krokhin and Yu.M.Popov, *Soviet JETP (U.S.S.R.)*, 40 (1961) 1897.
17. R.N.Hall, G.E.Fenner, J.D.Kingsley, T.J.Soltys and R.O.Carlson, *Phys.Rev.Letters*, 9 (1962) 366.
18. M.I.Nathan, W.P.Dumke, G.Burns, H.F.Dill and G.J.Lasher, *Appl.Phys.Letters*, 1 (1962) 62.
19. M.Bernard, personal communication.
20. A.M.Prochorov, *Soviet JETP (U.R.S.S.)*, 34 (1959) 1658.
21. A.L.Schawlow and C.H.Townes, *Phys.Rev.*, 112 (1958) 1940.
22. J.A.Armstrong and A.W Smith, *Appl.Phys.Letters*, 4 (1964) 196.
23. N.G.Basov, A.Z.Grasjuk and V.A.Katulin, *Dokl.Akad.Nauk (S.S.S.R.)*, 161 (1965) 1306.
24. C.Hilsum, *Lasers and their Applications*, London, 1964.
25. F.B.Alexander, *Appl.Phys.Letters*, 4 (1964) 13.
26. M.I.Nathan, *Proc.Electron.Elec.Engrs.*, 52 (1964) 770.
27. N.G.Basov and O.V.Bogdankevich, *Proc.Symp.Radiative Recombination in Semiconductors, Paris, 1964*, Dunod, Paris, 1965.

28. J. W. Crowe and R. W. Craig, *Appl. Phys. Letters*, 4 (1964) 57.
29. C. E. Hurwitz and R. J. Keyes, *Appl. Phys. Letters*, 5 (1964) 139.
30. V. S. Vavilov, E. L. Nolle and V. D. Egorov, *FTT (U. S. S. R.)*, 7 (1965) 934.
31. G. Burns, R. S. Levitt, M. I. Nathan and K. Weiser, *Proc. Electron. Elec. Engrs.*, 51 (1963) 1148.
32. T. Deutsch *et al.*, *Phys. Stat. Solids*, 3 (1963) 1001.
33. C. Benoit à la Guillaume and J. M. Debever, *Compt. Rend.*, 259 (1964) 2200.
34. I. Melngilis, *Appl. Phys. Letters*, 2 (1963) 176.
35. C. Benoit à la Guillaume and J. M. Debever, *Solid State Commun.*, 2 (1964) 145.
36. R. J. Phelan, A. R. Calawa, R. H. Rediker, R. J. Keyes and B. Lax, *Appl. Phys. Letters*, 3 (1963) 143.
37. N. Holonyak Jr. and S. F. Bevacqua, *Appl. Phys. Letters*, 1 (1962) 82.
38. T. M. Quist, R. H. Rediker, R. J. Keyes and W. E. Prag, *Bull. Am. Phys. Soc.*, (1963) 88.

Biography

Nikolai Gennadievich Basov was born on December, 14, 1922 in the small town of Usman near Voronezh, the son of Gennady Fedorovich Basov and Zinaida Andreevna Molchanova. His father was a professor of the Voronezh Forest Institute and devoted his life to investigation of the influence of forest belts on underground waters and on surface drainage.

After finishing secondary school in 1941 in Voronezh Basov was called up for military service and directed to the Kuibyshev Military Medical Academy. In 1943 he left the Academy with the qualification of a military doctor's assistant. He served in the Soviet Army and took part in the Second World War in the area of the First Ukranian Front. In December 1945, he was demobilized and entered the Moscow Institute of Physical Engineers where he studied theoretical and experimental physics.

From 1950 to 1953 he was a postgraduate student of the Moscow Institute of Physical Engineers. At that time, Basov was working on his thesis at the P. N. Lebedev Physical Institute of the Academy of Sciences, U. S. S. R., under the guidance of Professor M. A. Leontovich and Professor A. M. Prochorov.

In 1950 Basov joined the P. N. Lebedev Physical Institute, where at present he is vice-director and head of the laboratory of quantum radiophysics. He is also a professor of the department of solid-state physics at the Moscow Institute of Physical Engineers.

In 1952 Dr. Basov began to work in the field of quantum radiophysics. He made various attempts (firstly, theoretical and then experimental) to design and build oscillators (together with A. M. Prochorov). In 1956 he defended his doctoral thesis on the theme «A Molecular Oscillator», which summed up the theoretical and experimental works on creation of a molecular oscillator utilizing an ammonia beam.

In 1955 Basov organized a group for the investigation of the frequency stability of molecular oscillators. Together with his pupils and collaborators A. N. Oraevsky, V. V. Nikitin, G. M. Strakhovsky, V. S. Zuev and others, Dr. Basov studied the dependence of the oscillator frequency on different parameters for a series of ammonia spectral lines, proposed methods of increasing

the frequency stability by means of slowing down molecules, proposed methods of producing slow molecules, investigated the operation of oscillators with resonators in series, realized phase stabilization of klystron frequency by means of molecular oscillators, studied transition processes in molecular oscillators, and designed an oscillator utilizing a beam of deuterium ammonia. In the result of these investigations the oscillators with a frequency stability of 10^{-11} have been realized in 1962.

In 1957 Basov started to work on the design and construction of quantum oscillators in the optical range. A group of theorists and research workers began to study the possibilities for realization of quantum oscillators by means of semiconductors, and after A. Javan's proposal, the possibility of their realization in the gas media was also investigated. In 1958 together with B. M. Vul and Yu. M. Popov he investigated the conditions for production of states with a negative temperature in semiconductors, and suggested utilization of a pulse breakdown for that purpose. In 1961 together with O. N. Krokhin and Yu. M. Popov, Basov proposed three different methods for the obtaining of a negative temperature state in semiconductors in the presence of direct and indirect transitions (optical excitation, utilization of a beam of fast electrons and injection of carriers through a degenerated *p–n* junction).

As a result of a cooperative effort with B. M. Vul and collaborators the injection semiconductor lasers utilizing crystals of gallium arsenide were made at the beginning of 1963.

In 1964 semiconductor lasers with electronic excitation have been made (together with O. V. Bogdankevich and A. N. Devyatkov); and somewhat later, lasers with optical excitation were constructed (together with A. Z. Grasiuk and V. A. Katulin). For these achievements a group of scientists of Lebedev Physical Institute was awarded the Lenin Prize for 1964.

Beginning from 1961 Dr. Basov (together with V. S. Zuev, P. G. Kriukov, V. S. Letokhov *et al.*) carried out theoretical and experimental research in the field of powerful lasers. There have been found the ways of obtaining powerful short laser pulses. The nature of appearance of such pulses in quantum oscillators and their propagation in quantum amplifiers have been investigated. This work resulted in the development of high-power single-pulse Nd-glass lasers with 30 J energy and $2 \cdot 10^{-11}$ sec pulse duration (in 1968 together with P. G. Kriukov, Yu. V. Senatsky *et al.*) and multichannel lasers with energy 10^3 J within 10^{-9} sec (in 1971 in collaboration with G. V. Sklizkov *et al.*).

In 1962 N. Basov and O. N. Krokhin investigated the possibility of laser radiation usage for the obtaining of thermonuclear plasmas. In 1968 Basov

and his associates (P. G. Kriukov, Yu. V. Senatsky, S. D. Zakharov) have succeeded in observing for the first time neutron emission in the laser-produced deuterium plasmas. The spectra of multicharged ions CaXVI, FeXXIII, K XIX and others have also been observed (together with O. N. Krokhin, S. L. Mandelshtam, G. V. Sklizkov). There has been developed a theory of picosecond pulse formation (together with V. S. Letokhov). In the same year Basov and his associate A. N. Oraevsky proposed a method of the thermal laser excitation. Further theoretical considerations of this method by Basov, A. N. Oraevsky and V. A. Shegłov encouraged the development of the so-called gasdynamic lasers.

In 1963 Dr. Basov and his colleagues (V. V. Nikitin, Yu. M. Popov, V. N. Morozov) began to work in the field of optoelectronics. They developed in 1967 a number of fast-operating logic elements on the basis of diode lasers. At present a logic structure of the multichannel optoelectronic systems producing 10^{10} operations per second for the optical data processing is under the development.

The studies of the radiation of the condensed rare gases under the action of a powerful electron beam have been initiated in 1966 by Basov and his collaborators (V. A. Danilychev, Yu. M. Popov), and they were the first to obtain in 1970 the laser emission in the vacuum ultraviolet range.

In 1968 Basov (in cooperation with O. V. Bogdankevich and A. S. Nasibov) made a proposal for a laser projection TV. About the same time Dr. Basov (together with V. V. Nikitin) began the studies of the frequency standard in the optical range (on the basis of gas lasers). In 1970 they succeeded in realizing a gas laser stabilized in the methane absorption line with frequency stability 10^{-11}.

In 1969 Basov (together with E. M. Belenov and V. V. Nikitin) hypothesized that to obtain the frequency standard with the stability 10^{-12}–10^{-13} a ring laser can be used with a nonlinear absorption cell.

A large contribution has been made by Dr. Basov to the field of chemical lasers. In 1970 under his guidance an original chemical laser was achieved which operates on a mixture of deuterium, F and CO_2 at the atmospheric pressure. In the same year Basov (in cooperation with E. M. Belenov, V. A. Danilychev and A. F. Suchkov) proposed and developed experimentally an elion (electrical pumping of ionized compressed gases) method of gas laser excitation. Using this method for a CO_2 and N_2 mixture compressed to 25 atm., they achieved a great increase of power of the gas laser volume unit compared to the typical low pressure CO_2 lasers.

In the end of 1970 Basov (together with E. P. Markin, A. N. Oraevsky, A. V. Pankratov) presented experimental evidence for the stimulation of chemical reactions by the infrared laser radiation.

In 1959 Dr. Basov was awarded the Lenin Prize together with A. M. Prochorov for the investigation leading to the creation of molecular oscillators and paramagnetic amplifiers. In 1962 Dr. Basov was elected a corresponding member of the Academy of Sciences of the U. S. S. R.; in 1966, a member of the Academy; in 1967, a member of the Presidium of the Academy of Sciences of the U. S. S. R., and a foreign member of the German Academy of Sciences in Berlin; and in 1971, a foreign member of the German Academy «Leopoldina».

Dr. Basov is Editor-in-chief of the Soviet scientific journals «*Priroda*» (Nature) and «*Kvantovaya Elektornika*» (Quantum Electronics); he is also a member of the Editorial Board of «*Il Nuovo Cimento*».

In 1970 Dr. Basov was awarded the rank of Hero of Socialist Labour. Dr. Basov is a member of the Soviet Committee of the Defence of Peace and a member of the World Peace Council.

Nikolai Basov married in 1950. His wife, Ksenia Tikhonovna Basova, is also a physicist and is with the Department of General Physics of the Moscow Institute of Physical Engineers. They have two sons: Gennady (born 1954) and Dmitry (born 1963).

A. M. PROCHOROV

110

Quantum electronics

Nobel Lecture, December 11, 1964

One may assume as generally accepted that quantum electronics started to exist at the end of 1954–beginning of 1955[1,2]. Just by that time theoretical grounds had been created, and the first device,–a molecular oscillator–had been designed, and constructed. A basis for quantum electronics as a whole is the phenomenon of an induced radiation, predicted by A. Einstein in 1917. However, quantum electronics was developed considerably later. And it is quite natural to ask the questions: why did it happen so? What reasons put obstacles for the creation of quantum devices considerably earlier, for instance, in the period of 1930–1940?

In order to try to answer these questions I should like to say some words about the bases on which quantum electronics is founded.

As I have already noted, the phenomenon of an induced radiation was predicted by Einstein. It is well-known that an atom being in an excited state may give off its energy in the form of emission of radiation (quantum) in two ways. The first way is a spontaneous emission of radiation, *i.e.* when an atom emits energy without any external causation. All usual light sources (filament lamps, gas-discharge tubes, etc.) produce light by way of such spontaneous radiation. It means that scientist engaged in the field of optical spectroscopy were well acquainted with that type of emission many years ago.

The second way for an atom to give off its energy is through stimulated emission of radiation. That phenomenon was noted by Einstein to be necessary in order to describe thermodynamic equilibrium between an electromagnetic field and atoms. The phenomenon of stimulated emission occurs when an excited atom emits due to interaction with an external field (quantum). Then two quanta are involved: one is the external one, the other is emitted by the atom itself. Those two quanta are indistinguishable, *i.e.* their frequency and directivity coincide. This very significant characteristic of an induced radiation (which was, apparently, first pointed at by Dirac in 1927) made it possible to build quantum electronic devices.

In order to observe a stimulated emission, it is necessary, firstly, to have excited atoms and, secondly, that the probability of an induced radiation must

be greater than that of a spontaneous emission. If atoms are in a thermal equilibrium, optical levels are not populated. If the atoms became excited they make a transition to the lower level due to spontaneous emission. This happens because the probability of a stimulated emission radiation is small at usual densities of the light energy. Therefore scientists engaged in the field of spectroscopy did not take into account the stimulated radiation and some of them, apparently, considered that phenomenon as a «Kunststück» of a theorist necessary only for the theory.

It is absolutely clear that if all atoms are in an excited state, such a system of atoms will amplify the radiation, and many scientists understood this already before 1940, but none of them pointed to the possibility of creating light oscillators in this way. It may seem strange because, in principle, optical quantum oscillators (lasers) could have been made even before 1940. But definite fundamental results were necessary. They appeared after Second World War, when radiospectroscopy started to develop rapidly. And just the scientists engaged in the field of radiospectroscopy laid down foundations for quantum electronics[1,2].

How should one explain this? There were some favourable circumstances which had not been available to the scientists working in the field of optical spectroscopy.

First of all, since for systems in a thermal equilibrium, the excited levels in the radio range, contrary to the optical ones may have a large population and of course one should then take into account induced radiation. Indeed, if the concentration of particles on the lower level equals n_1, and on the excited level n_2, one may write down a net absorption coefficient for an electromagnetic wave in the form

$$\alpha = \frac{1}{v} h\nu (n_1 B_{12} - n_2 B_{21}) \tag{1}$$

the value of B_{12} characterizes the probability of an absorption act, and B_{21} characterizes the probability of an induced radiation act. If the levels are not degenerated, $B_{12} = B_{21}$, then will take the form

$$\alpha = \frac{1}{v} h\nu (n_1 - n_2) B_{12} \tag{2}$$

For a frequency as in the optical range, under usual conditions of thermal equilibrium, one may put with a high accuracy n_2 equal to zero, and then the absorption coefficient will become

$$\alpha = \frac{1}{v} h\nu \; n_1 \; B_{12} \tag{3}$$

Therefore, for the optical range the absorption coefficient depends only on the population of the lower level. For as in the radio range, as a rule, $h\nu << kT$. In that case

$$n_2 = n_1 e^{-\frac{h\nu}{kT}} \approx n_1 \left(1 - \frac{h\nu}{kT}\right)$$

Then the value of α will be

$$\alpha = \frac{1}{v} h\nu \; n_1 \; B_{12} \; \frac{h\nu}{kT} \tag{4}$$

As is seen from Eqn. 4, due to stimulated emission, the value of the absorption coefficient becomes reduced by a factor $kT/h\nu$ compared to what it would be without the presence of induced emission. Therefore, all scientists engaged in radiospectroscopy have to take into account the effect of induced radiation. Moreover, for increasing an absorption coefficient one has to lower the temperature in order to decrease the population of the upper level and to weaken, in this way, the influence of stimulated radiation. It follows from Eqn. 2 that for systems that are not in thermal equilibrium, but have $n_2 > n_1$, the net absorption coefficient becomes negative, *i.e.* such a system will amplify radiation. In principle, such systems were known to physicists long time ago for the radio range. If we pass molecular or atomic beams through inhomogeneous magnetic or electric fields, we can separate out molecules in definite state. In particular, one may obtain molecular beams containing molecules in the upper state only. Actually physicists engaged in the field of a micro-wave radiospectroscopy started to think about application of molecular beams for increasing the resolving power of radio spectroscopes. In order to gain a maximum absorption in such beams, one must have molecules either in the lower state only or in the upper states only, *i.e.* one must separate them using inhomogeneous electric or magnetic fields. If molecules are in the upper state, they will amplify a radiation.

As is well-known from radio engineering, any system able to amplify can be made to oscillate. For this purpose a feedback coupling is necessary. A theory for ordinary tube oscillators is well developed in the radio range. For description of those oscillators, the idea of a negative resistance or conductance is introduced, *i.e.* an element in which so-called negative losses take place. In

the case of a quantum oscillator the medium with a negative absorption factor is that «element». Therefore the condition of self-excitation for the quantum oscillator should be written in the similar way as for a tube oscillator. According to the analogy with usual tube oscillators, it is quite natural to expect that for a quantum oscillator the oscillations will also be quite monochromatic.

Finally, the resonator system is a very significant element of a quantum oscillator (maser or laser) as well as in any other oscillator with sinusoidal oscillations. However, resonator systems were well worked out for the radio range, and just those resonators operating in the radio range were used for masers. Thus, a very important element–a cavity–was also well-known to the scientists engaged in radiospectroscopy. Therefore all elements of masers really existed separately but it was necessary to do a very important step of synthesis in order to construct the maser. First two papers–one of which was published in the U.S.S.R. and the other in the U.S.A.–appeared independently; and they both were connected directly with the construction of radio spectroscopes with a high resolving power, using molecular beams. As is easily seen from the aforesaid, this result is quite natural.

Those two papers initiated the development of quantum electronics, and the first successes in this new field of physics stimulated its further progress. Already in 1955, there was proposed a new method–the method of pumping for gaining a negative absorption[3]. That method was further developed and applied for the construction of new types of quantum devices. In particular, the method of pumping was developed and applied for designing and building quantum amplifiers for the radio range on the basis of an electronic paramagnetic resonance[4,5]. Quantum devices according to the suggestion of Prof. Townes were called masers. One might think that after the successful construction of masers in the radio range, there would soon be made quantum oscillators (lasers) in the optical range as well. However, this did not occur. Those oscillators were constructed only 5–6 years later. What caused such a delay?

There were two difficulties. One of them was as follows: at that time no resonators for the optical wavelength range were available. The second difficulty was that no methods were immediately available for gaining an inverse population in the optical wavelength range.

Let us consider firstly the question of resonators. It is well-known that radio engineering started its development from the region of long waves where resonators were used in the form of self-inductance coils combined with condensers. In that case the size of the resonator is much less than one wavelength.

With development towards short waves the cavity resonators were used. They are closed volume cavities. The size of those cavities was comparable with a wavelength. It is quite clear that with the help of such cavities it is impossible to advance into the region of very short waves. In particular, it would be impossible to reach the optical range.

In 1958 there was proposed the so-called open type of cavities for masers and lasers in the region of very short waves[6,7]. Practically speaking this is Fabry–Perot's interferometer; however, a «radio engineering» approach made it possible to suggest using such a system as a resonator. Afterwards, spherical mirrors were used together with plane mirrors. The size of these resonators is much more than that wavelength.

At present open cavities are widely utilized for lasers.

There were also systems suggested for the production of a negative absorption in the submillimeter (far-infrared) wavelength range[6], the infrared and optical wave ranges[7–10]. Those works stimulated a further advancement in the region of shorter waves and, in particular, into the optical range. However, the first quantum-optical oscillator was made as late as 1960[11]. It was a ruby laser. After carrying out investigations in the optical range, many scientists started to think about further extension into the X-ray field. In that wavelength range the same difficulties arise as in the optical wavelength range. It was necessary to suggest new types of resonators and to find also the proper system that would produce negative absorption. As is known X-ray quantum oscillators have not yet been constructed. We have also considered this problem* and we have found that there are essential difficulties.

Indeed in the X-ray region the lifetime of an excited level state is small and one may assume that the line width is determined by that lifetime only. Then the absorption coefficient may be written in a very simple form

$$\alpha = \frac{\lambda^2}{4\pi}(n_1 - n_2) \tag{5}$$

where λ is the wavelength and n_1 and n_2 are the densities of particles in the lower and upper level stated respectively. As seen from Eqn. 5 the absorption coefficient decreases sharply as the wavelength becomes shorter. This is an extremely unpleasant circumstance. Indeed for the laser operation the value of α should be of the order of one inverse cm. If $\lambda = 1$ Å, the density of particles on the upper level must be not less than 10^{17} cm^{-3}. The lifetime in the upper

* This problem was also considered in ref. 1.

level τ is of the order of 10^{-16} sec. Therefore, 10^{33} particles/cm^3 per second must be excited. In order to fulfil this condition one has to overcome essential experimental difficulties.

Nevertheless, the successes of quantum electronics are enormous even without the construction of laser in the X-ray region.

At present the range in which lasers and masers operate is extremely wide. Recently a far-infrared range had not been available but now investigations in this region are carried out with a great success. In practice with the help of masers and lasers one may produce emission from the the lowest radio frequencies to the ultraviolet region.

Operation of all masers and lasers is based on the fact that in media with a negative absorption the processes of induced emission dominate due to a large field intensity over spontaneous or non-radiative transitions. Moreover, at present, for instance, one may produce with the help of a ruby laser such radiation energy densities at which the probability of multiquantum processes becomes comparable with the probability of one quantum process or even exceeds it. This is a new qualitative jump which leads to interesting results of several kinds.

First of all one may estimate[12] the maximum power which a ruby laser is able to give per cm^2. That power equals 10^{11} Watt/cm^2 that is one hundred gigaWatts/cm^2. At that power the probability of a simultaneous absorption of three quanta of red light with transition of an electron to the conduction band is so great that a further growth of the field stops. For three-quantum processes the losses grow in proportion to the cube of the energy density, *i.e.* a very strong dependence on the field takes place.

Large electric fields available in a laser beam may carry out ionization and dissociation of molecules and breakdown in a solid as well.

Multiquantum processes do not always have a bad effect (for instance, restriction of the maximum density given by laser) but they open up new possibilities for a further development of quantum electronics. This interesting and principally new direction is connected with the construction of lasers which utilize two-quantum transitions. It was pointed out in 1963 in the U.S.S.R.[13] and somewhat later but independently in the U.S.A.[14,15] that construction of these oscillators should be possible. The idea of this laser is that if there is an inverse population between two levels with the energy difference $E_2 - E_1 = h\nu$, generation of two frequencies ν_1 and ν_2 is possible in such a way that

$$\nu = \nu_1 + \nu_2 \tag{6}$$

In particular, frequencies ν_1 and ν_2 may coincide. However, frequencies ν_1 and ν_2 may have any value as long as only condition 6 is fulfilled.

Operation of such an oscillator, as it was mentioned above, is connected with two-quantum transitions, the probability of which is rather great, if the field density is considerable. For self-excitation of these oscillators it is necessary to have another oscillator of a sufficiently large initial energy density with frequencies ν_1 or ν_2, and one may remove the external field only after self-excitation of the two quantum oscillator. Such two quantum oscillators have two possibilities: (*1*) Faster growth of the field density than in the case of usual lasers. (*2*) Possibility of producing any frequency within the framework of the relation 6.

Construction of an oscillator for any given radiation frequency will greatly extend the region of application of lasers. It is clear that if we make a laser with a sweep frequency, we apparently shall be able to influence a molecule in such a way that definite bonds will be excited and, thus, chemical reactions will take place in certain directions.

However, this problem will not be simple even after design of the appropriate lasers. But one thing is clear: the problem is extremely interesting and perhaps its solution will be able to make a revolution in a series of branches of chemical industry.

1. N. G. Basov and A. M. Prochorov, *Zh. Eksperim. i Teor. Fiz.*, 27 (1954) 431.
2. J. P. Gordon, H. J. Zeiger and C. H. Townes, *Phys. Rev.*, 95 (1954) 282.
3. N. G. Basov and A. M. Prochorov, *Zh. Eksperim, i Teor. Fiz.*, 28 (1955) 249.
4. N. Bloembergen, *Phys. Rev.*, 104 (1956) 324.
5. H. E. D. Scovil, G. Feiher and H. Seidel, *Phys. Rev.*, 105 (1957) 762.
6. A. M. Prochorov, *Zh. Eksperim. i Teor. Fiz.*, 34 (1958) 1658.
7. A. L. Schawlow and C. H. Townes, *Phys. Rev.*, 112 (1958) 1940.
8. N. G. Basov, B. M. Vul and Yu. M. Popov, *Zh. Eksperim. i Teor. Fiz.*, 37 (1959) 587.
9. A. Javan, *Phys. Rev. Letters*, 3 (1959) 87.
10. F. A. Butaeva and V. A. Fabrikant, *O srede s otritsatelnym koeffitsentom pogloshcheniya. Issledovaniya po eksperimentalnoy i teoreticheskoy fizike.* Sbornik pamyati G. S. Landsberga, Izdatelstvo Akad. Nauk S. S. S. R., 1959, p. 62.
11. T. H. Maiman, *Brit. Commun. Electron.*, 1 (1960) 674.
12. F. V. Bunkin and A. M. Prochorov, in preparation.
13. A. M. Prochorov and A. C. Selivanenko, *U. S. S. R. Patent Application* No. 872,303, priority 24 Dec. 1963.
14. P. P. Sorokin and N. Braslau, *IBM J. Res. Develop.*, 8 (1964) 177.
15. R. Z. Gorwin, *IBM J. Res. Develop.*, 8 (1964) 338.

Biography

Alexander Mikhailovich Prochorov was born on July 11th, 1916, in Australia. After the Great October Revolution he went in 1923 with his parents to the Soviet Union.

In 1934 Alexander Prochorov entered the Physics Department of the Leningrad State University. He attended lectures of Prof. V. A. Fock (quantum mechanics, theory of relativity), Prof. S. E. Frish (general physics, spectroscopy), and Prof. E. K. Gross (molecular physics). After graduating in 1939 he became a postgraduate student of the P. N. Lebedev Physical Institute in Moscow, in the laboratory of oscillations headed by Academician N. D. Papaleksi. There he started to study the problems of propagation of radio waves. In June 1941, he was mobilized in the Red Army. He took part in the Second World War and was wounded twice. After his second injury in 1944, he was demobilized and went back to the laboratory of oscillations of the P. N. Lebedev Physical Institute. There he began to investigate nonlinear oscillations under the guidance of Prof. S. M. Rytov.

In 1946 he defended his thesis on the theme «Theory of Stabilization of Frequency of a Tube Oscillator in the Theory of a Small Parameter».

Starting in 1947, upon the suggestion of Academician V. I. Veksler, Prochorov carried out a study of the coherent radiation of electrons in the synchotron in the region of centimetre waves. As a result of these investigations he wrote and defended in 1951 his Ph. D. thesis «A Coherent Radiation of Electrons in the Synchotron Accelerator».

After the death of Academician I. D. Papaleksi in 1946, the laboratory of oscillations was headed by Academician M. A. Leontovich. Starting from 1950 being assistant chief of the laboratory, Prochorov began to investigate on a wide scale the question of radiospectroscopy and, somewhat later, of quantum electronics. He organized a group of young scientists interested in the subjects.

In 1954, when Academician M. A. Leontovich started to work in the Institute of Atomic Energy, Prochorov became head of the laboratory of oscillations, which position he still holds. In 1959 the laboratory of radio astronomy

headed by Prof. V. V. Vitkevitch) was organized from one of the departments of the laboratory of oscillations, and in 1962 another department was separated as the laboratory of quantum radiophysics (headed by Prof. N. G. Basov).

Academician D. V. Skobeltzyn, director of the Institute, and Academician M. A. Leontovich as well, rendered great assistance in the development of the research on radiospectroscopy and quantum electronics. The investigations carried out by Basov and Prochorov in the field of microwave spectroscopy resulted in the idea of a molecular oscillator. They developed theoretical grounds for creation of a molecular oscillator and also constructed a molecular oscillator operating on ammonia. In 1955, Basov and Prochorov proposed a method for the production of a negative absorption which was called the pumping method.

From 1950 to 1955, Prochorov and his collaborators carried out research on molecular structures by the methods of microwave spectroscopy.

In 1955 Professor Prochorov began to develop the research on electronic paramagnetic resonance (EPR). A cycle of investigations of EPR spectra and relaxation times in various crystals was carried out, in particular investigations on ions of the iron group elements in the lattice of Al_2O_3.

In 1955, Prochorov studied with A. A. Manenkov the EPR spectra of ruby that made it possible to suggest it as a material for lasers in 1957. They designed and constructed masers using various materials and studied characteristics of the masers as well. This research was done in cooperation with the laboratory of radiospectroscopy of the Institute of Nuclear Physics of the Moscow University; this laboratory was organized by Prochorov in 1957. One of the masers constructed for a wavelength of 21 cm is used in the investigations of the radioastronomical station of the Physical Institute in Pushino.

The EPR methods were also utilized for the study of free radicals. In particular, the transition of a free radical of DPPH from a paramagnetic state into an antiferromagnetic state at 0.3 °K was observed.

In 1958 Prochorov suggested a laser for generation of far-infrared waves. As a resonator it was proposed to use a new type of cavity which was later called «the cavity of an open type». Practically speaking, it is Fabri–Pero's interferometer. Similar cavities are widely used in lasers.

At present Prochorov's principal scientific interests lie in the field of solid lasers and their utilization for physical purposes, in particular for studies of multiquantum processes. In 1963, he suggested together with A. S. Selivanenko, a laser using two-quantum transitions.

Alexander Prochorov is Professor at the Moscow State University and Vice-President of URSI.

He married in 1941; his wife, G. A. Shelepina, is a geographer. They have one son.

121

Physics 1965

SIN-ITIRO TOMONAGA

JULIAN SCHWINGER

RICHARD P. FEYNMAN

«for their fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles»

Physics 1965

Presentation Speech by Professor Ivar Waller, member of the Nobel Committee for Physics

Your Majesty, Your Royal Highnesses, Ladies and Gentlemen.

The electrons of an atom move according to the laws of quantum mechanics established in 1925 and the next following years. For the hydrogen atom, which has only one electron and consequently is the simplest atom to investigate theoretically, the calculation of the motion of the electron in the electric field of the nucleus led to results of such accuracy that 20 years elapsed until any error of the theory could be found experimentally. This occurred, however, in 1947 when Lamb and his collaborator Retherford discovered that some energy levels of hydrogen which should coincide theoretically were in fact somewhat shifted relative to each other.

One important result of the work of this year's Nobel prize winners Sinitiro Tomonaga, Julian Schwinger and Richard Feynman was the explanation of the Lamb-shift. Their work is, however, much more general and of deep general significance to physics. It has explained and also predicted several important phenomena. It is the continuation of some investigations performed in the late 1920's in order to find the general quantum mechanical laws according to which the atoms and in particular the electrons give rise to electromagnetic fields, *e.g.* emit light, and are influenced by such fields. By applying quantum mechanics not only to matter but also to the electromagnetic field Dirac, Heisenberg, and Pauli managed in those years to formulate a theory, called quantum electrodynamics, which contains the quantum mechanical laws for the interaction of charged particles, in particular electrons, and the electromagnetic field. It satisfies the important condition of being in agreement with the theory or relativity.

It was soon realized, however, that this theory had serious defects. When one tried to calculate a quantity of such importance as the contribution to the mass of an electron originating in its interaction with the electromagnetic field an infinite and therefore useless result was obtained. A similar difficulty occurred for the charge of the electron.

Because of the fundamental importance of having a more useful quantum electrodynamics many theoretical physicists tried during the 1930's to over-

come those difficulties. Some indications were forthcoming how this should be accomplished. It lasted, however, until the 1940's for decisive progress to be made.

A new area was then initiated by investigations first performed by Tomonaga. His work was primarily related to the demands imposed by the theory of relativity. In a paper published in 1943 and in later work published together with his collaborators, Tomonaga managed to give a new formulation of quantum electrodynamics and other similar theories, which marked an important progress.

Definite progress was only made as a consequence of the discovery of the Lamb-shift mentioned earlier. When this discovery was discussed at a conference the idea was accepted that the new effect could be explained by quantum electrodynamics provided the proper interpretation was given to this theory. The correctness of this idea was supported by a provisional calculation of the Lamb-shift which was published by Bethe shortly after the conference.

As soon as Tomonaga knew about the Lamb experiment and Bethe's paper he realized that an essential step to be taken was to substitute the experimental mass for the fictive mechanical mass which appeared in the equations of quantum electrodynamics and to perform a similar renormalization of the electric charge. The compensating terms which had then to be introduced in the equations should cancel the infinities. Tomonaga managed to carry out this difficult program on the basis of his earlier investigations mentioned above. He deduced further a correct formula for the Lamb-shift which was found to give results in good agreement with the measurements.

Almost simultaneously with the discovery of the Lamb-shift another pecularity was found by Kusch and his collaborator Foley, which made it clear that the magnetic moment of the electron is somewhat larger than had been assumed before. Using the method of renormalization which he also developed Schwinger was able to prove that a small anomalous contribution should be added to the value of the magnetic moment accepted until then. His calculation agreed with the experiments. Schwinger's calculation was indeed earlier than and very important for the proper interpretation of these measurements.

Schwinger had developed the formalism of the new quantum electrodynamics in several fundamental papers using partially methods similar to those of Tomonaga. He has also made this formalism more useful for practical calculations.

Feynman used more radical methods for solving the problems of quantum electrodynamics. He created a new formalism which he made very useful for

practical calculations by introducing a graphical interpretation called Feynman diagrams, which have become an important feature of modern physics. In the description used by Feynman the electromagnetic field did not any more appear explicitly. His description has been very valuable in elementary particle physics where it is necessary to consider besides the electromagnetic also other interactions.

When considering the truth of quantum electrodynamics in its new form one has first of all to realize the extraordinary success of this theory in giving results in agreement with the experiments. For the Lamb-shift and for the anomalous part of the magnetic moment of the electron the agreement is within some parts in one hundred thousand respectively in a million and no disagreement has yet been found. Quantum electrodynamics is indeed one of the most accurate of all the theories of physics. Further evidence in this respect is given by the applications of the theory to the positronium atom and to the mu-particle. The new formalism has also been very important for other parts of physics in particular elementary-particle physics, but also solid-state physics, nuclear physics and statistical mechanics.

Professor Tomonaga has unfortunately been prevented by an accident from receiving his prize here in Stockholm. It will be presented to him by intermediary of the ambassador of Sweden in Tokyo, and it is accompanied by the congratulations of the Royal Academy of Science.

Professor Schwinger and Professor Feynman. By introducing new ideas and methods into an old theory you have, together with Professor Tomonaga, created a new and most successful quantum electrodynamics, which occupies a central position in physics. This theory has been unique in stimulating modern research. You have yourself contributed to the extension of its methods to other fields of physics where it has also been essential for recent progress.

On behalf of the Royal Academy of Science I congratulate you on your work and ask you to receive your Nobel prize from the hands of His Majesty the King.

SIN-ITIRO TOMONAGA

Development of quantum electrodynamics

Personal recollections

Nobel Lecture, May 6, 1966

(1) In 1932, when I started my research career as an assistant to Nishina, Dirac published a paper in the *Proceedings of the Royal Society, London*[1]. In this paper, he discussed the formulation of relativistic quantum mechanics, especially that of electrons interacting with the electromagnetic field. At that time a comprehensive theory of this interaction had been formally completed by Heisenberg and Pauli[2], but Dirac was not satisfied with this theory and tried to construct a new theory from a different point of view. Heisenberg and Pauli regarded the (electromagnetic) field itself as a dynamical system amenable to the Hamiltonian treatment; its interaction with particles could be described by an interaction energy, so that the usual method of Hamiltonian quantum mechanics could be applied. On the other hand, Dirac thought that the field and the particles should play essentially different roles. That is to say, according to him, «the role of the field is to provide a means for making observations of a system of particles» and therefore «we cannot suppose the field to be a dynamical system on the same footing as the particles and thus be something to be observed in the same way as the particles».

Based on such a philosophy, Dirac proposed a new theory, the so-called many-time theory, which, besides being a concrete example of his philosophy was of much more satisfactory and beautiful form than other theories presented up to then. In fact, from the relativistic point of view, these other theories had a common defect which was inherent in their Hamiltonian formalism. The Hamiltonian dynamics was developed on the basis of non-relativistic concepts which make a sharp distinction between time and space. It formulates a physical law by describing how the state of a dynamical system changes with time. Speaking quantum-mechanically, it is a formalism to describe how the probability amplitude changes with time t. Now, as an example, let us consider a system composed of N particles, and let the coordinates of each particle be $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N$. Then the probability amplitude of the system is a function of the N variables $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N$, and in addition, of the time t to which the amplitude is referred. Thus this function contains only one time

variable in contrast to N space variables. In the theory of relativity, however, time and space must be treated on an entirely equal footing so that the above imbalance is not satisfactory. On the other hand, in Dirac's theory which does not use the Hamiltonian formalism, it becomes possible to consider different time variables for each particle, so that the probability amplitude can be expressed as a function of $\boldsymbol{r}_1 t_1, \boldsymbol{r}_2 t_2, \ldots, \boldsymbol{r}_N t_N$. Accordingly, the theory satisfies the requirement of the principle of relativity that time and space be treated with complete equality. The reason why the theory is called the many-time theory is because N distinct time variables are used in this way.

This paper of Dirac's attracted my interest because of the novelty of its philosophy and the beauty of its form. Nishina also showed a great interest in this paper and suggested that I investigate the possibility of predicting some new phenomena by this theory. Then I started computations to see whether the Klein–Nishina formula could be derived from this theory or whether any modification of the formula might result. I found out immediately however, without performing the calculation through to the end, that it would yield the same answer as the previous theory. This new theory of Dirac's was in fact mathematically equivalent to the older Heisenberg–Pauli theory, and I realized during the calculation that one could pass from one to the other by a unitary transformation. The equivalence of these two theories was also discovered by Rosenfeld[3] and by Dirac–Fock–Podolsky[4] and was soon published in their papers.

Though Dirac's many-time formalism turned out to be equivalent to the Heisenberg–Pauli theory, it had the advantage that it gave us the possibility of generalizing the former interpretation of the probability amplitude. Namely, while one could calculate the probability of finding particles at points with coordinates $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N$, all at the time t according to the previous theory, one could now compute more generally the probability that the first particle is at $\boldsymbol{r}_1$ at time t_1, the second at $\boldsymbol{r}_2$ at time t_2, ..., and the N-th at $\boldsymbol{r}_N$ at time t_N. This was first discussed by Bloch[5] in 1934.

(*2*) In this many-time theory developed by Dirac, electrons were treated according to the particle picture. Alternatively, in quantum theory, any particle should be able to be treated according to the wave picture. As a matter of fact, electrons were also treated as waves in the Heisenberg–Pauli theory, and it was well known that this wave treatment was frequently more convenient than the particle treatment. So the question arose as to whether one could reformulate the Heisenberg–Pauli theory in a way which would be more

satisfactory relativistically, when electrons were treated as waves as well as the electromagnetic field.

As Dirac already pointed out, the Heisenberg–Pauli theory is built upon the Hamiltonian formalism and therefore the probability amplitude contains only one time variable. That is to say, the probability amplitude is given as a function of the field strength at different space points and of one common time variable. However, the concept of a common time at different space points does not have a relativistically covariant meaning.

Around 1942, Yukawa[6] wrote a paper emphasizing this unsatisfactory aspect of the quantum field theory. He thought it necessary to use the idea of the g.t.f. (generalized transformation function) proposed by Dirac[7] to correct this defect of the theory. Here I shall omit talking about the g.t.f., but, briefly, Yukawa's idea was to introduce as the basis of a new theory a concept which generalized the conventional conception of the probability amplitude. However, as pointed out also by Yukawa, we encounter the difficulty that, in doing this, cause and effect can not be clearly separated from each other. According to Yukawa, the inseparability of cause and effect would be an essential feature of quantum field theory, and without abandoning the causal way of thinking which strictly separates cause and effect, it would not be possible to solve various difficulties appearing in quantum field theory about which I will talk later. I thought however, that it might be possible (without introducing such a drastic change as Yukawa and Dirac tried to do) to remedy the unsatisfactory, unpleasant aspect of the Heisenberg–Pauli theory of having a common time at different space points. In other words, it should be possible, I thought, to define a relativistically meaningful probability amplitude which would be manifestly covariant, without being forced to give up the causal way of thinking. In having this expectation I was recalling Dirac's many-time theory which had enchanted me 10 years before.

When there are N particles in Dirac's many-time theory, we assign a time t_1 to the first particle, t_2 to the second, and so on, thus introducing N different times, $t_1, t_2, \ldots, t_N$, instead of the one common time t. Similarly, I tried in quantum field theory to see whether it was possible to assign different times, instead of one common time, to each space point. And in fact I was able to show that this was possible[8].

As there are an infinite number of space points in field theory in contrast to the finite number of particles in particle theory, the number of time variables appearing in the probability amplitude became infinite. But it turned out that no essential difficulty appeared. An interpretation quite analogous to the

one discussed by Bloch in connection with Dirac's many-time theory could be given to our probability amplitude containing an infinite number of time variables. Further, it was found that the theory thus formulated was completely covariant and that this covariant formulation was equivalent in its whole content to the Heisenberg–Pauli theory: it was shown, just as in the case of the many-time theory, that we could pass from one to the other by a unitary transformation. I began this work about 1942, and completed it in 1946.

(*3*) As I mentioned a little while ago, there are many difficulties in the quantum mechanics of fields. In particular, infinite quantities always arise which are associated with the presence of field reactions in various processes. The first phenomenon which attracted our attention as a manifestation of field reactions was the electromagnetic mass of the electron. The electron, having a charge, produces an electromagnetic field around itself. In turn, this field, the so-called self-field of the electron, interacts with the electron. We call this interaction the field reaction. Because of the field reaction the apparent mass of the electron differs from the original mass. The excess mass due to this field reaction is called the electromagnetic mass of the electron and the experimentally observed mass is the sum of the original mass and this electromagnetic mass. The concept of the electromagnetic mass had already appeared in the classical theory of the electron by Lorentz, who computed the electromagnetic mass by applying the classical theory and obtained the result that the mass becomes infinite for the point (zero size) electron. On the other hand, the electromagnetic mass was computed in quantum theory by various people, and here I mention particularly the work of Weisskopf[9]. According to him, the quantum-mechanical electromagnetic mass turned out to be infinite, and although the order of the divergence was much weaker than in the case of the Lorentz theory, the observed mass, which included this additional mass, would be infinite. This would be, of course, contrary to experiment.

In order to overcome the difficulty of an infinitely large electromagnetic mass, Lorentz considered the electron not to be point-like but to have a finite size. It is very difficult, however, to incorporate a finite sized electron into the framework of relativistic quantum theory. Many people tried various means to overcome this problem of infinite quanties, but nobody succeeded.

In connection with field reactions, the next problem which attracted the attention of physicists was determining what kind of influence the field reaction exerts in electron-scattering processes. Let us consider, as a concrete example, a problem in which an electron is scattered by an external field. In

the ordinary treatment, we neglect the effect of field reactions on the scattered electron, assuming that it is negligibly small. Then the behavior of the scattering obtained by calculation (*e.g.* the Rutherford formula) fits very well with experiment. But what will happen if the influence of field reaction is taken into account? This theoretical problem was examined non-relativistically by Braunbeck–Weinmann[10] and Pauli–Fierz[11] and relativistically by Dancoff[12].

While Dancoff applied an approximation method, the perturbation method, in his relativistic calculation, Pauli and Fierz treated the problem in such a way that the most important part of the field reaction was first separated out exactly by employing a contact transformation method which was similar to the one which Bloch–Nordsieck[13] had published a year before. Since Pauli and Fierz adopted a non-relativistic model, and further simplified the problem by using the so-called dipole approximation, their calculation was especially transparent. At any rate, both non-relativistic and relativistic calculations exhibited several infinities in the scattering processes*.

The conclusions of these people were fatal to the theory. That is, the influence of the field reaction becomes infinite in this problem. The effect of field reaction on a quantity called the scattering cross section, which expresses quantitatively the behavior of the scattering, rather than becoming negligibly small, becomes infinitely large. This does not, of course, agree with experiment.

This discouraging state of affairs generated in many people a strong distrust of quantum field theory. There were even those with the extreme view that the concept of field reaction itself had nothing to do at all with the true law of nature.

On the other hand, there was also the view that the field reaction might not be altogether meaningless but would play an essential role in the scattering processes, though the appearance of divergences revealed a defect of the theory. Heisenberg[14], in his paper published in 1949, emphasized that the field reaction would be crucial in meson–nucleon scattering. Just at that time I was studying at Leipzig, and I still remember vividly how Heisenberg enthusiastically explained this idea to me and handed me galley proofs of his forthcoming paper. Influenced by Heisenberg, I came to believe that the

* The main purpose of the work of Bloch–Nordsieck and Pauli–Fierz was to solve the so-called infrared catastrophe which was one of a number of divergences. Since this difficulty was resolved in their papers we confine ourselves here to a discussion of the other divergences which are of the so-called ultraviolet type.

problem of field reactions far from being meaningless was one which required a frontal attack.

Thus, after coming back to Japan from Leipzig, I began to examine the nature of the infinities appearing in scattering processes at the same time that I was engaged in the above-mentioned work of formulating a covariant field theory. What I wanted to know was what kind of relationship exists between the infinity associated with the scattering process and that associated with the mass. If you read the above-mentioned papers of Bloch–Nordsieck and Pauli–Fierz, you will see that one of the terms containing infinite quantities is first separated out by a contact transformation and this term turns out to be just the term modifying the mass. Besides this kind of infinity there appeared, according to Pauli–Fierz, another kind of infinity characteristic of the scattering process. I further investigated a couple of simple models which were not realistic, but could be solved exactly. What was understood from these models, was that the most strongly divergent terms in the scattering process had the same form as the expression giving the modification of the particle mass due to field reactions, and therefore both should be manifestations of the same effect. In other words, at least a portion of the infinities appearing in the scattering process could be amalgamated into the infinity associated with the particle mass, leaving infinities proper to the scattering process alone. These turned out to be more weakly divergent than the infinity associated with the mass.

Since these conclusions were derived from non-relativistic or unrealistic models, it was still doubtful whether the same thing would occur in the case of relativistic electrons interacting with the electromagnetic field. Dancoff tried to answer this question. He calculated relativistically the infinities appearing in the scattering process and determined which of them could be amalgamated into the mass and which remained as infinities proper to the scattering process alone. He found that there remained, in the latter group of infinite terms, one which was at least as divergent as the infinity of the mass, a finding which differed from the conclusion based on fictitious models.

Actually, there are two kinds of field reactions in the case of the relativistic electron and electromagnetic field. One of them ought to be called «of mass type» and the other «of vacuum polarization type». The field reaction of mass type changes the apparent electronic mass from its original value by the amount of the electromagnetic mass as was calculated by Weisskopf. On the other hand, the field reaction of vacuum polarization type changes the apparent electronic charge from its original value. As was discussed in further papers

by Weisskopf[15] and others, infinite terms appear in the apparent electronic charge if the effect of vacuum polarization is taken into account. However, in this talk, for simplicity, I will mention only briefly the divergence of the vacuum polarization type.

(*4*) In the meantime, in 1946, Sakata[16] proposed a promising method of eliminating the divergence of the electron mass by introducing the idea of a field of cohesive force. It was the idea that there exists unknown field, of the type of the meson field which interacts with the electron in addition to the electromagnetic field. Sakata named this field the cohesive force field, because the apparent electronic mass due to the interaction of this field and the electron, though infinite, is negative and therefore the existence of this field could stabilize the electron in some sense. Sakata pointed out the possibility that the electromagnetic mass and the negative new mass cancel each other and that the infinity could be eliminated by suitably choosing the coupling constant between this field and the electron. Thus the difficulty which had troubled people for a long time seemed to disappear insofar as the mass was concerned. (It was found later that Pais[17] proposed the same idea in the U. S. independently of Sakata.) Then what concerned me most was whether the infinities appearing in the electron-scattering process could also be removed by the idea of a plus–minus cancellation.

An example of a computation of how the field reaction influences the scattering process was already given by Dancoff. What we had to do was just to replace the electromagnetic field by the cohesive force field in Dancoff's calculation. I mobilized young people around me and we performed the computation together[18]. Infinities with negative sign actually appeared in the scattering cross-section as was expected. However, when we compared these with the infinities with positive sign which Dancoff calculated for the electromagnetic field, the two infinities did not cancel each other completely. That is to say, according to our result, the Sakata theory led to the cancellation of infinities for the mass but not for the scattering process. It was also known that the infinity of vacuum polarization type was not cancelled by the introduction of the cohesive force field.

Unfortunately, Dancoff did not publish the detailed calculations in his paper, and while we were engaged in the above considerations, we felt it necessary to do Dancoff's calculation over again for ourselves in parallel with the computation of the influence of the cohesive force field. At the same time I happened to discover a simpler method of calculation.

This new method of calculation was to use the technique of contact transformations based on the previously mentioned formalism of the covariant field theory and was in a sense a relativistic generalization of the Pauli–Fierz method. This method had the advantage of separating the electromagnetic mass from the beginning, just as was shown in their paper.

Our new method of calculation was not at all different in its contents from Dancoff's perturbation method, but had the advantage of making the calculation more clear. In fact, what took a few months in the Dancoff type of calculation could be done in a few weeks. And it was by this method that a mistake was discovered in Dancoff's calculation; we had also made the same mistake in the beginning. Owing to this new, more lucid method, we noticed that, among the various terms appearing in both Dancoff's and our previous calculation, one term had been overlooked. There was only one missing term, but it was crucial to the final conclusion. Indeed, if we corrected this eror, the infinities appearing in the scattering process of an electron due to the electromagnetic and cohesive force fields cancelled completely, except for the divergence of vacuum polarization type.

(5) When this unfortunate error of Dancoff's was discovered, we had to reexamine his conclusions concerning the relation between the divergence of the scattering process and the divergence of the mass, in particular, the conclusion that there remained a portion of the infinities of the scattering process which could not be amalgamated into the modification of the mass. In fact, it turned out that after correcting the error, the infinity of mass type appearing in the scattering process could be reduced completely to the modification of the mass, and the remaining field reaction belonging to the scattering proper was not divergent[19]. In other words, the highest divergence part of the infinities appearing in the scattering process, in the relativistic as well as in the non-relativistic case, could be attributed to the infinity of mass. The reason why the remaining part became finite in the relativistic case was due to the fact that the order of the highest divergence was only $\log \infty$, and after amalgamating the divergence into the mass term, the remainder was convergent. The great value of this method of contact transformations was that once the infinity of the mass was separated out, we obtained a divergence-free theoretical framework.

In this way the nature of various infinities became fairly clear. Though I did not describe here the infinity of vacuum polarization type, this too appears in the scattering process, as mentioned earlier. However, Dancoff had already

discovered that this infinity could be amalgamated into an apparent change in the electronic charge. To state the conclusion, therefore, all infinities appearing in the scattering process can be attributed either to the infinity of the electromagnetic mass or to the infinity appearing in the electronic charge—there are no other divergences in the theory.

It is a very pleasant thing that no divergence is involved in the theory except for the two infinities of the electronic mass and charge. We cannot say that we have no divergences in the theory, since the mass and charge are in fact infinite. It is to be noticed, however, that if we reduce the infinities appearing in the scattering process to modifications of mass and charge, the remaining terms all become finite. Further, if we examine the structure of the theory, after the infinities are amalgamated into the mass and charge terms, we see that the only mass and charge appearing in the theory are the values modified by field reactions – the original values and excess ones due to field reactions never appear separately.

This situation gives rise to the following possibility. The theory does not of course yield a resolution of the infinities. That is, since those parts of the modified mass and charge due to field reactions contain divergence, it is impossible to calculate them by the theory. However, the mass and charge observed in experiments are not the original mass and charge but the mass and charge as modified by field reactions, and they are finite. On the other hand, the mass and charge appearing in the theory are, as I mentioned above, after all the values modified by field reactions. Since this is so, and particularly since the theory is unable to calculate the modified mass and charge, we may adopt the procedure of substituting experimental values for them phenomenologically. When a theory is incompetent in part, it is a common procedure to rely on experiment for that part. This procedure is called the renormalization of mass and charge, and our method has brought the possibility that the theory will lead to finite results by the renormalization even if it contains defects.

The idea of renormalization is far from new. Many people used explicitly or implicitly this idea, and we find the word renormalization already in Dancoff's paper. In his calculation it appeared, because of an error that there still remained a divergence in the scattering even after the renormalization of the electron mass. This error was very unfortunate; if he had performed the calculation correctly, the history of renormalization theory would have been completely different.

(*6*) This period, around 1946–1948, was soon after the second world war, and

it was quite difficult in Japan to obtain information from abroad. But soon we got the news that in the U. S., Lewis and Epstein[20] found Dancoff's mistake and gave the same conclusions as ours, Schwinger[21] constructed a covariant field theory similar to ours, and he was probably performing various calculations making use of it. In particular, little by little news arrived that the so-called Lamb-shift was discovered[22] as a manifestation of the electromagnetic field reaction and that Bethe[23] was calculating it theoretically. The first information concerning the Lamb-shift was obtained not through the *Physical Review*, but through the popular science column of a weekly U. S. magazine. This information about the Lamb-shift prompted us to begin a calculation more exactly than Bethe's tentative one.

The Lamb-shift is a phenomenon in which the energy levels of a hydrogen atom show some shifts from the levels given by the Dirac theory. Bethe thought that the field reactions were primarily responsible for this shift. According to his calculation, field reactions give rise to an infinite level shift, but he thought that it should be possible to make it finite by a mass renormalization and a tentative calculation yielded a value almost in agreement with experiments.

This problem of the level shift is different from the scattering process, but it was conceivable that the renormalization which was effective in avoiding infinities in the scattering process would be workable in this case as well. In fact, the contact transformation method of Pauli and Fierz devised to solve the scattering problem could be applied to this case, clarifying Bethe's calculation and justifying his idea. Therefore the method of covariant contact transformations, by which we did Dancoff's calculation over again would also be useful for the problem of performing the relativistic calculation for the Lamb-shift. This was our prediction.

The calculation of the Lamb-shift was done by many people in the U. S.[24]. Among others, Schwinger, commanding powerful mathematical techniques, and by making thorough use of the method of covariant contact transformations, very skilfully calculated not only the Lamb-shift but other quantities such as the anomalous magnetic moment of the electron. After long, laborious calculations, less skilful than Schwinger's, we[25] obtained a result for the Lamb-shift which was in agreement with Americans'. Furthermore, Feynman[26] devised a convenient method based on an ingenious idea which could be used to extend the approximation of Schwinger and ours to higher orders, and Dyson[27] showed that all infinities appearing in quantum electrodynamics could be treated by the renormalization procedure to an arbitrarily high order of approximation. Furthermore, this method devised by Feynman and de-

veloped by Dyson was shown by many people to be applicable not only to quantum electrodynamics, but to statistical mechanics and solid-state physics as well, and provided a new, powerful method in these fields. However, these matters will probably be discussed by Schwinger and Feynman themselves and need not be explained by me. So far I have told you the story of how I played a tiny, partial role in the recent development of quantum electrodynamics, and here I would like to end my talk.

1. Dirac, *Proc. Roy. Soc. (London)*, 136 (1932) 453.
2. Heisenberg and Pauli, *Z. Physik*, 56 (1929) 1.
3. Rosenfeld, *Z. Physik*, 76 (1932) 729.
4. Dirac, Fock and Podolsky, *Phys. U. S. S. R.*, 2 (1932) 468.
5. Bloch, *Phys. Z. U. S. S. R.*, 5 (1943) 301.
6. Yukawa, *Kagaku (Tokyo)*, 12 (1943) 249.
7. Dirac, *Phys. Z. U. S. S. R.*, 3 (1933) 64.
8. Tomonaga, *Progr. Theoret. Phys. (Kyoto)*, 1 (1946) 27; Koba, Tati and Tomonaga, *ibid.*, 2 (1947) 101, 198; Kanesawa and Tomonaga, *ibid.*, 3 (1948) 1, 101.
9. Weisskopf, *Phys. Rev.*, 56 (1939) 72.
10. Braunbeck and Weinman, *Z. Phys.*, 110 (1938) 369.
11. Pauli and Fierz, *Nuovo Cimento*, 15 (1938) 267.
12. Dancoff, *Phys. Rev.*, 55 (1939) 959.
13. Bloch and Nordsieck, *Phys. Rev.*, 52 (1937) 54.
14. Heisenberg, *Z. Phys.*, 113 (1939) 61.
15. Weisskopf, *Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd.*, 14 (1936) No. 6.
16. Sakata, *Progr. Theoret. Phys. (Kyoto)*, 1 (1946) 143.
17. Pais, *Phys. Rev.*, 73 (1946) 173.
18. Ito, Koba and Tomonaga, *Progr. Theoret. Phys. (Kyoto)*, 3 (1948) 276; Koba and Takeda, *ibid.*, 3 (1948) 407.
19. Koba and Tomonaga, *Progr. Theoret. Phys. (Kyoto)*, 3 (1948) 290; Tati and Tomonaga, *ibid.*, 3 (1948) 391.
20. Lewis, *Phys. Rev.*, 73 (1948) 173; Epstein, *ibid.*, 73 (1948) 179.
21. Schwinger, *Phys. Rev.*, 73 (1948) 416, 74 (1948) 1439, 75 (1949) 651, 76 (1949) 790.
22. Lamb and Rutherford, *Phys. Rev.*, 72 (1947) 241.
23. Bethe, *Phys. Rev.*, 72 (1947) 339.
24. Kroll and Lamb, *Phys. Rev.*, 75 (1949) 388; French and Weisskopf, *ibid.*, 75 (1949) 1241.
25. Fukuda, Miyamoto and Tomonaga, *Progr. Theoret. Phys. (Kyoto)*, 4 (1949) 47, 121.
26. Feynman, *Phys. Rev.*, 74 (1948) 1430, 76 (1949) 769.
27. Dyson, *Phys. Rev.*, 75 (1949) 486, 1736.

Biography 137

Sin-itiro Tomonaga was born in Tokyo, Japan, on March 31, 1906, the eldest son of Sanjuro Tomonaga and Hide Tomonaga. In 1913 his family moved to Kyoto when his father was appointed a professor of philosophy at Kyoto Imperial University. From that time he was brought up in Kyoto. He is a graduate of the Third Higher School, Kyoto, a renowned senior high school which has educated a number of leading personalities in prewar Japan.

Tomonaga completed work for Rigakushi (bachelor's degree) in physics at Kyoto Imperial University in 1929, with one of his intimate friends. Dr. Hideki Yukawa, Nobel laureate. He was engaged in graduate work for three years at the same university and was then appointed a research associate by Dr. Yoshio Nishina at the Institute of Physical and Chemical Research, Tokyo, where he started to work in a newly developed frontier of theoretical physics – quantum electrodynamics – under the guidance of Dr. Nishina. His paper on the photoelectric pair creation is well-known.

Tomonaga stayed in Leipzig, Germany, from 1937 to 1939, to study nuclear physics and the quantum field theory in collaboration with the theoretical group of Dr. W. Heisenberg, where he published a paper «Innere Reibung und Wärmeleitfähigkeit der Kernmaterie», which was chosen as the thesis for Rigakuhakushi (Doctor of Science) at Tokyo Imperial University in December, 1939.

In 1940, Dr. Tomonaga directed his attention to the meson theory and developed the intermediate coupling theory in order to clarify the structure of the meson cloud around the nucleon. He joined the faculty of Tokyo Bunrika University (which was absorbed into the Tokyo University of Education in 1949) as Professor of Physics in 1941. It was in 1942 when he first proposed the covariant formulation of the quantum field theory in which the concept of the quantum state was generalized so as to be relativistically covariant.

During the Second World War, Dr. Tomonaga was interested in developing a theory of microwave systems. He solved the motion of electrons in the magnetron and also developed a unified theory of the systems consisting of wave guides and cavity resonators.

As soon as the War was over, Tomonaga came back to academic research again with a programme in which he was first to summarize and extend the intermediate coupling theory and secondly to apply the covariant field theory to actual physical systems. His aim was to investigate the nature of field reaction in the meson theory as well as in quantum electrodynamics. He was confident, prior to the Lamb–Rutherford experiment, by means of a model calculation that divergence difficulty in quantum electrodynamics could be overcome simply by handling the infinite mass and charge due to field reactions in some way or another. It was only a step further for him to develop the renormalization theory with covariant formalism in his right hand and experimental support in his left.

Dr. Tomonaga was invited to the Institute for Advanced Study, Princeton, in 1949 where he was engaged in the investigation of a one-dimensional fermion system. Thus he first succeeded in clarifying the nature of collective oscillations of a quantum-mechanical many-body system and opened a new frontier of theoretical physics, modern many-body problem. In 1955, he published an elementary theory of quantum mechanical collective motions.

Dr. Tomonaga took the leadership in establishing the Institute for Nuclear Study, University of Tokyo, in 1955. From 1956 to 1962 he was appointed President of the Tokyo University of Education and since 1963 he has been President of the Science Council of Japan and Director of the Institute for Optical Research, Tokyo University of Education. He occupies an important position in various governmental committees for scientific research and policy making.

Tomonaga's honours and awards include the Japan Academy Prize (1948); the Order of Culture (1952); the Lomonosov Medal, U. S. S. R. (1964).

Dr. Tomonaga is a member of the Japan Academy, the Deutsche Akademie der Naturforscher «Leopoldina» and a foreign member of the Royal Swedish Academy of Science. He is a corresponding member of the Bayerische Akademie der Wissenschaften and a foreign associate of the National Academy of Science of U. S. A.

Tomonaga has published widely in scientific journals on such subjects as quantum electrodynamics, the meson theory, nuclear physics, cosmic rays, and the many-body problem. His book, «Quantum Mechanics», was published in 1949 and translated into English in 1963.

Tomonaga was married in 1940 to Ryoko Sekiguchi, daughter of Dr. K. Sekiguchi, the former Director of the Tokyo Metropolitan Observatory.

They have two sons, Atsushi and Makoto and one daughter. Their daughter was married in 1965 to Dr. Y. Nagashima, research associate of the Physics Department, University of Rochester.

JULIAN SCHWINGER

Relativistic quantum field theory

Nobel Lecture, December 11, 1965

The relativistic quantum theory of fields was born some thirty-five years ago through the paternal efforts of Dirac, Heisenberg, Pauli and others. It was a somewhat retarded youngster, however, and first reached adolescence seventeen years later, an event which we are gathered here to celebrate. But it is the subsequent development and more mature phase of the subject that I wish to discuss briefly today.

I shall begin by describing to you the logical foundations of relativistic quantum field theory. No dry recital of lifeless «axioms» is intended but, rather, an outline of its organic growth and development as the synthesis of quantum mechanics with relativity. Indeed, relativistic quantum mechanics – the union of the complementarity principle of Bohr with the relativity principle of Einstein – *is* quantum field theory. I beg your indulgence for the mode of expression I must often use. Mathematics is the natural language of theoretical physics. It is the irreplaceable instrument for the penetration of realms of physical phenomena far beyond the ordinary experience upon which conventional language is based.

Improvements in the formal presentation of quantum mechanical principles, utilizing the concept of action, have been interesting by-products of work in quantum field theory. Both my efforts in this direction[1] and those of Feynman[2] (which began earlier) were based on a study of Dirac concerning the correspondence between the quantum transformation function and the classical action. We followed quite different paths, however, and two distinct formulations of quantum mechanics emerged which can be distinguished as differential and integral viewpoints.

In order to suggest the conceptual advantages of these formulations, I shall indicate how the differential version transcends the correspondence principle and incorporates, on the same footing, two different kinds of quantum dynamical variable. It is just these two types that are demanded empirically by the two known varieties of particle statistics. The familiar properties of the variables $q_k, p_k, k = 1 \cdots n$, of the conventional quantum system enable one to *derive* the form of the quantum action principle. It is a differential statement about

time transformation functions,

$$\delta\langle t_1|t_2\rangle = (i/\hbar)\langle t_1|\delta[\textstyle\int_{t_2}^{t_1} dtL]|t_2\rangle \tag{1}$$

which is valid for a certain class of kinematical and dynamical variations. The quantum Lagrangian operator of this system can be given the very symmetrical form

$$L = \sum_{k=1}^{n} \frac{1}{4}\left(p_k\frac{dq_k}{dt} - q_k\frac{dp_k}{dt} + \frac{dq_k}{dt}p_k - \frac{dp_k}{dt}q_k\right) - H(q,p,t) \tag{2}$$

The symmetry is emphasized by collecting all the variables into the $2n$-component Hermitian vector $z(t)$ and writing

$$L = \frac{1}{4}\left(za\frac{dz}{dt} - \frac{dz}{dt}az\right) - H(z,t) \tag{3}$$

where a is a real antisymmetrical matrix, which only connects the complementary pairs of variables.

The transformation function depends explicitly upon the choice of terminal states and implicitly upon the dynamical nature of the system. If the latter is held fixed, any alteration of the transformation function must refer to changes in the states, as given by

$$\delta\langle t_1| = (i/\hbar)\langle t_1|G_1 \qquad \delta|t_2\rangle = -(i/\hbar)\,G_2|t_2\rangle \tag{4}$$

where G_1 and G_2 are infinitesimal Hermitian operators constructed from dynamical variables of the system at the specified times. For a given dynamical system, then,

$$\delta[\textstyle\int_{t_2}^{t_1} dtL] = G_1 - G_2 \tag{5}$$

which is the quantum principle of stationary action, or Hamilton's principle, since there is no reference on the right hand side to variations at intermediate times. The stationary action principle implies equations of motion for the dynamical variables and supplies explicit expressions for the infinitesimal operators $G_{1,2}$. The interpretation of these operators as generators of transformations on states, and on the dynamical variables, implies commutation relations. In this way, all quantum-dynamical aspects of the system are derived from a single dynamical principle. The specific form of the commutation relations obtained from the symmetrical treatment of the usual quantum system is given by the matrix statement

$$[z(t), z(t)] = i\hbar a^{-1} \tag{6}$$

Note particularly how the antisymmetry of the commutator matches the antisymmetry of the matrix a.

We may now ask whether this general form of Lagrangian operator,

$$L = \frac{1}{4}\left(xA\frac{dx}{dt} - \frac{dx}{dt}Ax\right) - H(x,t) \tag{7}$$

also describes other kinds of quantum systems, if the properties of the matrix A and of the Hermitian variables x are not initially assigned. There is one general restriction on the matrix A, however. It must be skew-Hermitian, as in the realization by the real, antisymmetrical matrix a. Only one other simple possibility then appears, that of an imaginary, symmetrical matrix. We write that kind of matrix as $i\alpha$, where α is real and symmetrical, and designate the corresponding variables collectively by $\zeta(t)$. The replacement of the antisymmetrical a by the symmetrical α requires that the antisymmetrical commutators which characterize $z(t)$ be replaced by symmetrical anticommutators for $\varsigma(t)$, and indeed

$$\{\zeta(t), \zeta(t)\} = \hbar\,\alpha^{-1} \tag{8}$$

specifies the quantum nature of this second class of quantum variable. It has no classical analogue. The consistency of various aspects of the formalism requires only that the Lagrangian operator be an even function of this second type of quantum variable.

Time appears in quantum mechanics as a continuous parameter which represents an abstraction of the dynamical role of the measurement apparatus. The requirement of relativistic invariance invites the extension of this abstraction to include space and time coordinates. The implication that space-time localized measurements are a useful, if practically unrealizable idealization may be incorrect, but it is a grave error to dismiss the concept on the basis of *a priori* notions of measurability. Microscopic measurement has no meaning apart from a theory, and the idealized measurement concepts that are implicit in a particular theory must be accepted or rejected in accordance with the final success or failure of that theory to fulfill its avowed aims. Quantum field theory has failed no significant test, nor can any decisive confrontation be anticipated in the near future.

Classical mechanics is a determinate theory. Knowledge of the state at a given time permits precise prediction of the result of measuring any property of the system. In contrast, quantum mechanics is only statistically determinate. It is the probability of attaining a particular result on measuring any property

of the system, not the outcome of an individual microscopic observation, that is predictable from knowledge of the state. But both theories are causal – a knowledge of the state at one time implies knowledge of the state at a later time. A quantum state is specified by particular values of an optimum set of compatible physical properties, which are in number related to the number of degrees of freedom of the system. In a relativistic theory, the concepts of «before» and «after» have no intrinsic meaning for regions that are in space-like relation. This implies that measurements individually associated with different regions in space-like relation are causally independent, or compatible. Such measurements can be combined in the complete specification of a state. But since there is no limit to the number of disjoint spatial regions that can be considered, a relativistic quantum system has an infinite number of degrees of freedom.

The latter statement, incidentally, contains an implicit appeal to a general property that the mathematics of physical theories must possess – the mathematical description of nature is not sensitive to modifications in physically irrelevant details. An infinite total spatial volume is an idealization of the finite volume defined by the macroscopic measurement apparatus. Arbitrarily small volume elements are idealizations of cells with linear dimensions far below the level of some least distance that is physically significant. Thus, it would be more accurate, conceptually, to assert that a relativistic quantum system has a number of degrees of freedom that is extravagantly large, but finite.

The distinctive features of relativistic quantum mechanics flow from the idea that each small element of three-dimensional space at a given time is physically independent of all other such volume elements. Let us label the various degrees of freedom explicitly – by a point of three-dimensional space (in a limiting sense), and by other quantities of finite multiplicity. The dynamical variables then appear as

$$\chi_{a,\mathbf{x}}(t) \equiv \chi_a(t = x^0, \mathbf{x}) \tag{9}$$

which are a finite number of Hermitian operator functions of space–time coordinates, or quantum fields. The dynamical independence of the individual volume elements is expressed by a corresponding additivity of the Lagrangian operator

$$L = \int (\mathrm{d}\mathbf{x})\, \mathscr{L} \tag{10}$$

where the Lagrange function $\mathscr{L}$ describes the dynamical situation in the in-

finitesimal neighborhood of a point. The characteristic time derivative or kinematical part of L appears analogously in $\mathscr{L}$ in terms of the variables associated with the specified spatial point. The relativistic structure of the action principle is completed by demanding that it present the same form, independently of the particular partitioning of space–time into space and time. This is facilitated by the appearance of the action operator, the time integral of the Lagrangian, as the space–time integral of the Lagrange function. Accordingly, we require, as a sufficient condition, that the latter be a scalar function of its field variables, which implies that the known form of the time derivative term is supplemented by similar space derivative contributions. This is conveyed by

$$\mathscr{L} = \frac{1}{4}\left(\chi A^\mu \partial_\mu \chi - \partial_\mu \chi A^\mu \chi\right) - \mathscr{H}(\chi) \tag{11}$$

where the A^μ are a set of four finite skew-Hermitian matrices. A specific physical field is associated with submatrices of the A^μ, which are real and antisymmetrical for a field φ that obeys Bose–Einstein statistics, or imaginary and symmetrical for a field ψ obeying Fermi–Dirac statistics. Finally, the boundaries of the four-dimensional integration region, formed by three-dimensional space at the terminal times, are described by the invariant concept of the space-like surface σ, a three-dimensional manifold such that every pair of points is in space-like relation. The ensuing invariant form of the action principle of relativistic quantum field theory is (we now use atomic units, in which $\hbar = c = 1$)

$$\delta\langle\sigma_1|\sigma_2\rangle = i\langle\sigma_1|\delta\left[\textstyle\int_{\sigma_2}^{\sigma_1}(\mathrm{d}x)\,\mathscr{L}\right]|\sigma_2\rangle \tag{12}$$

Relativity is a statement of equivalence within a class of descriptions associated with similar but different measurement apparatus. Space–time coordinates are an abstraction of the role that the measurement apparatus plays in defining a space–time frame of reference. The empirical fact, that all connected space–time locations and orientations of the measurement apparatus supply equivalent descriptions, is interpreted by the mathematical requirement of invariance under the group of proper orthochronous inhomogenous Lorentz transformations, applied to the continuous numerical coordinates. There is another numerical element in the quantum-mechanical description that has a measure of arbitrariness and expresses an aspect of relativity. I am referring to the quantum-mechanical use of complex numbers and of the mathematical equivalence of the two square roots of -1, $\pm i$. What general property of

any measurement apparatus is subject to our control, in principle, but offers only the choice of two alternatives? The answer is clear—a macroscopic material system can be constructed of matter, or of antimatter! But let us not conclude too hastily that a matter apparatus and an antimatter apparatus are completely equivalent. It is characteristic of quantum mechanics that the dividing line between apparatus and system under investigation can be drawn somewhat arbitrarily, as long as the measurement apparatus always possesses the classical aspects required for the unambiguous recording of an observation. To preserve this feature, the interchange of matter and antimatter must be made on the whole assemblage of macroscopic apparatus and microscopic system. Since the observational label of this duality is the algebraic sign of electric charge, the microscopic interchange must reverse the vector of electric current j^{μ}, while maintaining the tensor $T^{\mu\nu}$ that gives the flux of energy and momentum. But this is just the effect of the coordinate transformation that reflects all four coordinates.

It is indeed true that the action principle does not retain its general form under either of the two transformations, the replacement of i with $-i$, and the reflection of all coordinates, but does preserve it under their combined influence. In more detail, the effect of complex conjugation is equivalent to the reversal of operator multiplication, which distinguishes fields with the two types of statistics. The reflection of all coordinates, a proper transformation, can be generated by rotations in the attached Euclidean space obtained by introducing the imaginary time coordinate $x_4 = i\, x^0$. This transformation alters reality properties, distinguishing fields with integral and half-integral spin. The combination of the two transformations replaces the original Lagrange function

$$\mathscr{L}(\varphi_{\text{int}}, \varphi_{1/2\,\text{int}}, \psi_{\text{int}}, \psi_{1/2\,\text{int}}) \tag{13}$$

with

$$\mathscr{L}(\varphi_{\text{int}}, i\varphi_{1/2\,\text{int}}, i\psi_{\text{int}}, \psi_{1/2\,\text{int}}) \tag{14}$$

If only fields of the types φ_{int}, $\Psi_{\pm\,\text{int}}$ are considered, which is the empirical connection between spin and statistics, the action principle is unaltered in form. This invariance property of the action principle[3] expresses the relativity of matter and antimatter. That is the content of the so-called TCP theorem. The anomalous response of the field types $\varphi_{\pm\,\text{int}}$, Ψ_{int} is also the basis for the theoretical rejection of these possibilities as contrary to general physical requirements of positiveness, namely, the positiveness of probability, and the positiveness of energy.

The concept of space-like surface is not limited to plane surfaces. According to the action principle, an infinitesimal deformation of the space-like surface on which a state is specified changes that state by

$$\delta\langle\sigma| = i\langle\sigma|\int d\sigma_\mu \, T^{\mu\nu}\delta x^\nu \tag{15}$$

which is the infinitely multiple relativistic generalization of the Schrödinger equation

$$\delta\langle t| = i\langle t|H(-\delta t) \tag{16}$$

This set of differential equations must obey integrability conditions, which are commutator statements about the elements of the tensor $T^{\mu\nu}$. Since rigid displacements and rotations can be produced from arbitrary local deformations, the operator expressions of the group properties of Lorentz transformations must be a consequence of these commutator conditions. Foremost among the latter are the equal-time commutators of the energy density T^{00}, which suffice to convey all aspects of relativistic invariance that are not of a three-dimensional nature. A system that is invariant under three-dimensional translations and rotations will be Lorentz invariant if, at equal times[4],

$$-i[T^{00}(x), T^{00}(x')] = -(T^{0k}(x) + T^{0k}(x'))\,\partial_k\,\delta(\mathbf{x}-\mathbf{x}') \tag{17}$$

This is a sufficient condition. Additional terms with higher derivatives of the delta function will occur, in general. But there is a distinguished class of physical system, which I shall call local, for which no further term appears. The phrase «local system» can be given a physical definition within the framework we have used or, alternatively, by viewing the commutator condition as a measurability statement about the property involved in the response of a system to a weak external gravitational field[5]. Only the external gravitational potential g_{00} is relevant here. A physical system is local if the operators $T^{\mu\nu}$, which may be explicit functions of g_{00} at the same time, do not depend upon time derivatives of g_{00}. The class of local systems is limited[6] to fields of spin 0, ½, 1. Such fields are distinguished by their physical simplicity in comparison with fields of higher spin. One may even question whether consistent relativistic quantum field theories can be constructed for non-local systems.

The energy density commutator condition is a very useful test of relativistic invariance. Only a month or so ago I employed it to examine whether a relativistic quantum field theory could be devised to describe magnetic as well as electric charge. Dirac pointed out many years ago that the existence of magnetic charge would imply a quantization of electric charge, in the sense that

the product of two elementary charges, $eg/\hbar c$, could assume only certain values. According to Dirac, these values are any integer or half-integer. In recent years, the theoretical possibility of magnetic charge has been attacked from several directions. The most serious accusation is that the concept is in violation of Lorentz invariance. This is sometimes expressed in the language of field theory by the remark that no manifestly scalar Lagrange function can be constructed for a system composed of electromagnetic field and electric and magnetic charge-bearing fields. Now it is true that there is no relativistically invariant theory for arbitary e and g, so that no formally invariant version could exist. Indeed, the unnecessary assumption that $\mathscr{L}$ is a scalar must be relinquished in favor of the more general possibilities that are compatible with the action principle. But the energy commutator condition can still be applied. I have been able to show that energy and momentum density operators can be exhibited which satisfy the commutator condition, together with the three-dimensional requirements, provided $eg/\hbar c$ possesses one of a discrete set of values. These values are integers, which is more restrictive than Dirac's quantization condition. Such general considerations shed no light on the empirical elusiveness of magnetic charge. They only emphasize that this novel theoretical possibility should not be dismissed lightly.

The physical systems that obey the commutator statement of locality do not include the gravitational field. But, this field, like the electromagnetic field, requires very special consideration. And these considerations make full use of the relativistic field concept. The dynamics of the electromagnetic field is characterized by invariance under gauge transformations, in which the phase of every charge-bearing field is altered arbitrarily, but continuously, at each space-time point while electromagnetic potentials are transformed inhomogeneously. The introduction of the gravitational field involves, not only the use of general coordinates and coordinate transformations, but the establishment at each point of an independent Lorentz frame. The gravitational field gauge transformations are produced by the arbitrary reorientation of these local coordinate systems at each point while gravitational potentials are transformed linearly and inhomogeneously. The formal extension of the action principle to include the gravitational field can be carried out[7], together with the verification of consistency conditions analogous to the energy density commutator condition[8]. To appreciate this tour de force, one must realize that the operator in the role of energy density is a function of the gravitational field, which is influenced by the energy density. Thus the object to be tested is only known implicitly. It also appears that the detailed specification of the

spatial distribution of energy lacks physical significance when gravitational phenomena are important. Only integral quantities or equivalent asymptotic field properties are physically meaningful in that circumstance. It is in the further study of such boundary conditions that one may hope to comprehend the significance of the gravitational field as the physical mediator between the worlds of the microscopic and the macroscopic, the atom and the galaxy.

I have now spoken at some length about fields. But it is in the language of particles that observational material is presented. How are these concepts related? Let us turn for a moment to the early history of our subject. The quantized field appears initially as a device for describing arbitrary numbers of indistinguishable particles. It was defined as the creator or annihilator of a particle at the specified point of space and time. This picture changed somewhat as a consequence of the developments in quantum electrodynamics to which Feynman, Tomonaga, myself, and many others contributed. It began to be appreciated that the observed properties of so-called elementary particles are partly determined by the effect of interactions. The fields used in the dynamical description were then associated with noninteracting or bare particles, but there was still a direct correspondence with physical particles. The weakness of electromagnetic interactions, as measured by the small value of the fine structure constant $e^2/\hbar c$ is relevant here, for the same view-point failed disastrously when extended to strongly interacting nucleons and mesons. The resulting wide spread disillusionment with quantum field theory is an unhappy chapter in the history of high-energy theoretical physics, although it did serve to direct attention toward various useful phenomenological calculation techniques.

The great qualitative difference between weakly interacting and strongly interacting systems was impressed upon me by a particular consideration which I shall now sketch for you[9]. In the absence of interactions there is an immediate connection between the quantized Maxwell field and a physical particle of zero mass, the photon. The null mass of the photon is the particle transcription of a field property, that electromagnetism has no well-defined range but weakens geometrically. Now one of the most important interaction aspects of quantum electrodynamics is the phenomenon of vacuum polarization. A variable electromagnetic field induces secondary currents, even in the absence of actual particle creation. In particular, a localized charge creates a counter charge in its vicinity, which partially neutralizes the effect of the given charge at large distances. The implication that physical charges are weaker than bare charges by a universal factor is the basis for

charge renormalization. But once the idea of a partial neutralization of charge is admitted one cannot exclude the possibility of total charge neutralization. This will occur if the interaction exceeds a certain strength such that an oppositely charged particle combination, of the same nature as the photon, becomes so tightly bound that the corresponding mass diminishes to zero. Under these circumstances no long-range fields would remain and the massless particle does not exist. We learn that the connection between the Maxwell field and the photon is not an *a priori* one, but involves a specific dynamical aspect, that electromagnetic interactions are weaker than the critical strength. It is a natural speculation that another such field exists which couples more strongly than the critical amount to nucleonic charge, the property carried by all heavy fermions. That hypothesis would explain the absolute stability of the proton, in analogy with the electromagnetic explanation of electron stability, without challenging the uniqueness of the photon.

A field operator is a localized excitation which, applied to the vacuum state, generates all possible energy-momentum, or equivalently, mass states that share the other distinguishing properties of the field. The products of field operators widen and ultimately exhaust the various classes of mass states. If an isolated mass value occurs in a particular product, the state is that of a stable particle with corresponding characteristics. Should a small neighborhood of a particular mass be emphasized, the situation is that of an unstable particle, with a proper lifetime which varies inversely as the mass width of the excitation. The quantitative properties of the stable and unstable particles that may be implied by a given dynamical field theory cannot be predicted with presently available calculation techniques. In these matters, to borrow a phrase of Ingmar Bergman, and St. Paul, we see through a glass, darkly. Yet, in the plausible qualitative inference that a substantial number of particles, stable and unstable, will exist for sufficiently strong interactions among a few fields lies the great promise of relativistic quantum field theory.

Experiment reveals an ever growing number and variety of unstable particles, which seem to differ in no essential way from the stable and long-lived particles with which they are grouped in tentative classification schemes. Surely one must hope that this bewildering complexity is the dynamical manifestation of a conceptually simpler substratum, which need not be directly meaningful on the observational level of particles. The relativistic field concept is a specific realization of this general groping toward a new conception of matter.

There is empirical evidence in favor of such simplification at a deeper dy-

namical level. Strongly interacting particles have been rather successfully classified with the aid of a particular internal symmetry group. It is the unitary group SU_3. The dimensionalities of particle multiplets that have been identified thus far are 1, 8, and 10. But the fundamental multiplet of dimensionality 3 is missing. It is difficult to believe in the physical significance of some transformation group without admitting the existence of objects that respond to the transformations of that group. Accordingly, I would describe the observed situation as follows. There are sets of fundamental fields that form triplets[10] with respect to the group U_3. The excitations produced by these fields are very massive and highly unstable. The low lying mass excitations of mesons and baryons are generated by products of the fundamental fields. If these fields are assigned spin $\frac{1}{2}$, as a specific model, it is sufficient to consider certain products of two and three fields to represent the general properties of mesons and baryons, respectively.

The cogency of this picture is emphasized by its role in clarifying a recent development in symmetry classification schemes. That is the provocative but somewhat mysterious suggestion that the internal symmetry group SU_3 be combined with space–time spin transformations to form the larger unitary group SU_6. This idea, with its relativistic generalizations, has had some striking numerical successes but there are severe conceptual problems in reconciling Lorentz invariance with any union of internal and space–time transformations, as long as one insists on immediate particle interpretation. The situation is different if one can refer to the space–time localizability that is the hallmark of the field concept[11]. Let us assume that the interactions among the fundamental fields are of such strength that field products at practically coincident points suffice to describe the excitation of the known relatively low lying particles. The resulting quasi-local structures are in some sense fields that are associated with the physical particles. I call these phenomenological fields, as distinguished from the fundamental fields which are the basic dynamical variables of the system. Linear transformations on the fundamental fields can simulate the effect of external probes, which may involve both unitary and spin degrees of freedom. If these external perturbations are sufficiently gentle, the structure of the particles will be maintained and the phenomenological field transformed linearly with indefinite multiplets. It is not implausible that the highly localized interactions among the phenomenological fields will exhibit a corresponding symmetry. Thus, combined spin and unitary transformations appear as a device for characterizing some gross features of the unknown inner field dynamics of physical particles, as it operates in the

neighborhood of a specific point. But these transformations can have no general significance for the transfer of excitations from point to point, and only lesser symmetries will survive in the final particle description.

Phenomenological fields are the basic concept in formulating the practical calculation methods of strong interaction field theory. They serve to isolate the formidable problem of the dynamical origin of physical particles from the more immediate questions referring to their properties and interactions. In somewhat analogous circumstances, those of non-relativistic many-particle physics, the methods and viewpoint of quantum field theory[12] have been enormously successful. They have clarified the whole range of cooperative phenomena, while employing relatively simple approximation schemes. I believe that phenomenological relativistic quantum field theory has a similar future, and will replace the algorithms that were introduced during the period of revolt from field theory. But the intuition that serves so well in non-relativistic contexts does not exist for these new conditions. One has still to appreciate the precise rules of phenomenological relativistic field theory, which must supply a self-consistent description of the residual interactions, given that the strong fundamental interactions have operated to compose the various physical particles. And when this is done, how much shall we have learned, and how much will remain unknown, about the mechanism that builds matter from more primitive constituents? Are we not at this moment,

> *... like stout Cortez when with eagle eyes*
> *He star'd at the Pacific—and all his men*
> *Look'd at each other with a wild surmise—*
> *Silent, upon a peak in Darien.*

And now it only remains for me to say: Tack så mycket för uppmärksamheten.

1. Some references are: J. Schwinger, *Phys. Rev.*, 82 (1951) 914, 91 (1953) 713; *Proc. Natl. Acad. Sci. (U. S.)*, 46 (1960) 883.
2. R. Feynman, *Rev. Mod. Phys.*, 20 (1948) 36; R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York, 1965.
3. In the first two papers cited in Ref. 1 I have assumed space-reflection invariance and shown the equivalence between the spin-statistics relation and the invariance of the action principle under combined time reflection and complex conjugation. It was later remarked by Pauli that the separate hypothesis of space-reflection invariance

was unnecessary, W. Pauli, *Niels Bohr and the Development of Physics*, McGraw-Hill, New York, 1955.
4. J. Schwinger, *Phys. Rev.*, 127 (1962) 324.
5. J. Schwinger, *Phys. Rev.*, 130 (1963) 406.
6. J. Schwinger, *Phys. Rev.*, 130 (1963) 800.
7. J. Schwinger, *Phys. Rev.*, 130 (1963) 1253.
8. J. Schwinger, *Phys. Rev.*, 132 (1963) 1317.
9. J. Schwinger, *Phys. Rev.*, 125 (1962) 397; 128 (1962) 2425.
10. J. Schwinger, *Phys. Rev.*, 135 (1964) B 816; 136 (1964) B 1821.
11. J. Schwinger, *Second Coral Gables Conference on Symmetry Principles at High Energy*, Freeman, San Francisco, 1965; *Phys. Rev.*, 140 (1965) B 158.
12. The general theory is described by P. Martin and J. Schwinger, *Phys. Rev.*, 115 (1959) 1342.

Biography

Julian Schwinger was born on 12th February 1918 in New York City. The principal direction of his life was fixed at an early age by an intense awareness of physics, and its study became an all-engrossing activity. To judge by a first publication, he debuted as a professional physicist at the age of sixteen. He was allowed to progress rapidly through the public school system of New York City. Through the kind interest of some friends, and especially I. I. Rabi of Columbia University, he transferred to that institution, where he completed his college education. Although his thesis had been written some two or three years earlier, it was in 1939 that he received the Ph. D. degree.

For the next two years he was at the University of California, Berkeley, first as a National Research Fellow and then as assistant to J. R. Oppenheimer. The outbreak of the Pacific war found Schwinger as an Instructor, teaching elementary physics to engineering students at Purdue University.

War activities were largely confined to the Radiation Laboratory at the Massachusetts Institute of Technology in Cambridge. Being a confirmed solitary worker, he became the night research staff. More scientific influences were also at work. He first approached electromagnetic radar problems as a nuclear physicist, but soon began to think of nuclear physics in the language of electrical engineering. That would eventually emerge as the effective range formulation of nuclear scattering. Then, being conscious of the large microwave powers available, Schwinger began to think about electron accelerators, which led to the question of radiation by electrons in magnetic fields. In studying the latter problem he was reminded, at the classical level, that the reaction of the electron's field alters the properties of the particle, including its mass. This would be significant in the intensive developments of quantum electrodynamics, which were soon to follow.

With the termination of the war Dr. Schwinger accepted an appointment as Associate Professor at Harvard University. Two years later he became full Professor. That was also the year of his marriage to Clarice Carrol of Boston.

In subsequent years, he worked in a number of directions, but there was a pattern of concentration on general theoretical questions rather than specific

problems of immediate experimental concern, which were nearer to the center of his earlier work. A speculative approach to physics has its dangers, but it can have its rewards. Schwinger was particularly pleased by an anticipation, early in 1957, of the existence of two different neutrinos associated, respectively, with the electron and the muon. This has been confirmed experimentally only rather recently. A related and somewhat earlier speculation, that all weak interactions are transmitted by heavy, charged, unit-spin particles still awaits a decisive experimental test. Schwinger's policy of finding theoretical virtues in experimentally unknown particles has culminated recently in a revived concern with magnetically charged particles, which may also be involved in the understanding of strong interactions.

In later years, Schwinger has followed his own advice about the practical importance of a phenomenological theory of particles. He has invented and systematically developed source theory, which deals uniformly with strongly interacting particles, photons, and gravitons, thus providing a general approach to all physical phenomena. This work has been described in two volumes published under the title «*Particles, Sources, and Fields*».

Awards and other honors include the first Einstein Prize (1951), the U.S. National Medal of Science (1964), honorary D.Sc. degrees from Purdue University (1961) and Harvard University (1962), and the Nature of Light Award of the U.S. National Academy of Sciences (1949). Prof. Schwinger is a member of the latter body, and a sponsor of the *Bulletin of the Atomic Scientists*.

RICHARD P. FEYNMAN

The development of the space-time view of quantum electrodynamics

Nobel Lecture, December 11, 1965

We have a habit in writing articles published in scientific journals to make the work as finished as possible, to cover all the tracks, to not worry about the blind alleys or to describe how you had the wrong idea first, and so on. So there isn't any place to publish, in a dignified manner, what you actually did in order to get to do the work, although, there has been in these days, some interest in this kind of thing. Since winning the prize is a personal thing, I thought I could be excused in this particular situation, if I were to talk personally about my relationship to quantum electrodynamics, rather than to discuss the subject itself in a refined and finished fashion. Furthermore, since there are three people who have won the prize in physics, if they are all going to be talking about quantum electrodynamics itself, one might become bored with the subject. So, what I would like to tell you about today are the sequence of events, really the sequence of ideas, which occurred, and by which I finally came out the other end with an unsolved problem for which I ultimately received a prize.

I realize that a truly scientific paper would be of greater value, but such a paper I could publish in regular journals. So, I shall use this Nobel Lecture as an opportunity to do something of less value, but which I cannot do elsewhere. I ask your indulgence in another manner. I shall include details of anecdotes which are of no value either scientifically, nor for understanding the development of ideas. They are included only to make the lecture more entertaining.

I worked on this problem about eight years until the final publication in 1947. The beginning of the thing was at the Massachusetts Institute of Technology, when I was an undergraduate student reading about the known physics, learning slowly about all these things that people were worrying about, and realizing ultimately that the fundamental problem of the day was that the quantum theory of electricity and magnetism was not completely satisfactory. This I gathered from books like those of Heitler and Dirac. I was inspired by the remarks in these books; not by the parts in which everything was proved and demonstrated carefully and calculated, because I couldn't

understand those very well. At the young age what I could understand were the remarks about the fact that this doesn't make any sense, and the last sentence of the book of Dirac I can still remember, «It seems that some essentially new physical ideas are here needed.» So, I had this as a challenge and an inspiration. I also had a personal feeling, that since they didn't get a satisfactory answer to the problem I wanted to solve, I don't have to pay a lot of attention to what they did do.

I did gather from my readings, however, that two things were the source of the difficulties with the quantum electrodynamical theories. The first was an infinite energy of interaction of the electron with itself. And this difficulty existed even in the classical theory. The other difficulty came from some infinites which had to do with the infinite numbers of degrees of freedom in the field. As I understood it at the time (as nearly as I can remember) this was simply the difficulty that if you quantized the harmonic oscillators of the field (say in a box) each oscillator has a ground state energy of $(1/2)\hbar\omega$ and there is an infinite number of modes in a box of every increasing frequency ω, and therefore there is an infinite energy in the box. I now realize that that wasn't a completely correct statement of the central problem; it can be removed simply by changing the zero from which energy is measured. At any rate, I believed that the difficulty arose somehow from a combination of the electron acting on itself and the infinite number of degrees of freedom of the field.

Well, it seemed to me quite evident that the idea that a particle acts on itself, that the electrical force acts on the same particle that generates it, is not a necessary one–it is a sort of a silly one, as a matter of fact. And, so I suggested to myself, that electrons cannot act on themselves, they can only act on other electrons. That means there is no field at all. You see, if all charges contribute to making a single common field, and if that common field acts back on all the charges, then each charge must act back on itself. Well, that was where the mistake was, there was no field. It was just that when you shook one charge, another would shake later. There was a direct interaction between charges, albeit with a delay. The law of force connecting the motion of one charge with another would just involve a delay. Shake this one, that one shakes later. The sun atom shakes; my eye electron shakes eight minutes later, because of a direct interaction across.

Now, this has the attractive feature that it solves both problems at once. First, I can say immediately, I don't let the electron act on itself, I just let this act on that, hence, no self-energy! Secondly, there is not an infinite number of degrees of freedom in the field. There is no field at all; or if you insist on

thinking in terms of ideas like that of a field, this field is always completely determined by the action of the particles which produce it. You shake this particle, it shakes that one, but if you want to think in a field way, the field, if it's there, would be entirely determined by the matter which generates it, and therefore, the field does not have any *independent* degrees of freedom and the infinities from the degrees of freedom would then be removed. As a matter of fact, when we look out anywhere and see light, we can always «see» some matter as the source of the light. We don't just see light (except recently some radio reception has been found with no apparent material source).

You see then that my general plan was to first solve the classical problem, to get rid of the infinite self-energies in the classical theory, and to hope that when I made a quantum theory of it, everything would just be fine.

That was the beginning, and the idea seemed so obvious to me and so elegant that I fell deeply in love with it. And, like falling in love with a woman, it is only possible if you do not know much about her, so you cannot see her faults. The faults will become apparent later, but after the love is strong enough to hold you to her. So, I was held to this theory, in spite of all difficulties, by my youthful enthusiasm.

Then I went to graduate school and somewhere along the line I learned what was wrong with the idea that an electron does not act on itself. When you accelerate an electron it radiates energy and you have to do extra work to account for that energy. The extra force against which this work is done is called the force of radiation resistance. The origin of this extra force was identified in those days, following Lorentz, as the action of the electron itself. The first term of this action, of the electron on itself, gave a kind of inertia (not quite relativistically satisfactory). But that inertia-like term was infinite for a point-charge. Yet the next term in the sequence gave an energy loss rate, which for a point-charge agrees exactly with the rate you get by calculating how much energy is radiated. So, the force of radiation resistance, which is absolutely necessary for the conservation of energy would disappear if I said that a charge could not act on itself.

So, I learned in the interim when I went to graduate school the glaringly obvious fault of my own theory. But, I was still in love with the original theory, and was still thinking that with it lay the solution to the difficulties of quantum electrodynamics. So, I continued to try on and off to save it somehow. I must have some action develop on a given electron when I accelerate it to account for radiation resistance. But, if I let electrons only act on other electrons the only possible source for this action is another electron in the

world. So, one day, when I was working for Professor Wheeler and could no longer solve the problem that he had given me, I thought about this again and I calculated the following. Suppose I have two charges–I shake the first charge, which I think of as a source and this makes the second one shake, but the second one shaking produces an effect back on the source. And so, I calculated how much that effect back on the first charge was, hoping it might add up the force of radiation resistance. It didn't come out right, of course, but I went to Professor Wheeler and told him my ideas. He said, –yes, but the answer you get for the problem with the two charges that you just mentioned will, unfortunately, depend upon the charge and the mass of the second charge and will vary inversely as the square of the distance R, between the charges, while the force of radiation resistance depends on none of these things. I thought, surely, he had computed it himself, but now having become a professor, I know that one can be wise enough to see immediately what some graduate student takes several weeks to develop. He also pointed out something that also bothered me, that if we had a situation with many charges all around the original source at roughly uniform density and if we added the effect of all the surrounding charges the inverse R square would be compensated by the R^2 in the volume element and we would get a result proportional to the thickness of the layer, which would go to infinity. That is, one would have an infinite total effect back at the source. And, finally he said to me, and you forgot something else, when you accelerate the first charge, the second acts later, and then the reaction back here at the source would be still later. In other words, the action occurs at the wrong time. I suddenly realized what a stupid fellow I am, for what I had described and calculated was just ordinary reflected light, not radiation reaction.

But, as I was stupid, so was Professor Wheeler that much more clever. For he then went on to give a lecture as though he had worked this all out before and was completely prepared, but he had not, he worked it out as he went along. First, he said, let us suppose that the return action by the charges in the absorber reaches the source by advanced waves as well as by the ordinary retarded waves of reflected light; so that the law of interaction acts backward in time, as well as forward in time. I was enough of a physicist at that time not to say, «Oh, no, how could that be?» For today all physicists know from studying Einstein and Bohr, that sometimes an idea which looks completely paradoxical at first, if analyzed to completion in all detail and in experimental situations, may, in fact, not be paradoxical. So, it did not bother me any more

than it bothered Professor Wheeler to use advance waves for the back reaction –a solution of Maxwell's equations, which previously had not been physically used.

Professor Wheeler used advanced waves to get the reaction back at the right time and then he suggested this: If there were lots of electrons in the absorber, there would be an index of refraction n, so, the retarded waves coming from the source would have their wave lengths slightly modified in going through the absorber. Now, if we shall assume that the advanced waves come back from the absorber without an index–why? I don't know, let's assume they come back without an index–then, there will be a gradual shifting in phase between the return and the original signal so that we would only have to figure that the contributions act as if they come from only a finite thickness, that of the first wave zone. (More specifically, up to that depth where the phase in the medium is shifted appreciably from what it would be in vacuum, a thickness proportional to $\lambda/(n-1)$.) Now, the less the number of electrons in here, the less each contributes, but the thicker will be the layer that effectively contributes because with less electrons, the index differs less from 1. The higher the charges of these electrons, the more each contribute, but the thinner the effective layer, because the index would be higher. And when we estimated it, (calculated without being careful to keep the correct numerical factor) sure enough, it came out that the action back at the source was completely independent of the properties of the charges that were in the surrounding absorber. Further, it was of just the right character to represent radiation resistance, but we were unable to see if it was just exactly the right size. He sent me home with orders to figure out exactly how much advanced and how much retarded wave we need to get the thing to come out numerically right, and after that, figure out what happens to the advanced effects that you would expect if you put a test charge here close to the source? For if all charges generate advanced, as well as retarded effects, why would that test not be affected by the advanced waves from the source?

I found that you get the right answer if you use half-advanced and half-retarded as the field generated by each charge. That is, one is to use the solution of Maxwell's equation which is symmetrical in time and that the reason we got no advanced effects at a point close to the source in spite of the fact that the source was producing an advanced field is this. Suppose the source s surrounded by a spherical absorbing wall ten light seconds away, and that the test charge is one second to the right of the source. Then the source is as much

as eleven seconds away from some parts of the wall and only nine seconds away from other parts. The source acting at time $t=0$ induces motions in the wall at time $+10$. Advanced effects from this can act on the test charge as early as eleven seconds earlier, or at $t=-1$. This is just at the time that the direct advanced waves from the source should reach the test charge, and it turns out the two effects are exactly equal and opposite and cancel out! At the later time $+1$ effects on the test charge from the source and from the walls are again equal, but this time are of the same sign and add to convert the half-retarded wave of the source to full retarded strength.

Thus, it became clear that there was the possibility that if we assume all actions are *via* half-advanced and half-retarded solutions of Maxwell's equations and assume that all sources are surrounded by material absorbing all the the light which is emitted, then we could account for radiation resistance as a direct action of the charges of the absorber acting back by advanced waves on the source.

Many months were devoted to checking all these points. I worked to show that everything is independent of the shape of the container, and so on, that the laws are exactly right, and that the advanced effects really cancel in every case. We always tried to increase the efficiency of our demonstrations, and to see with more and more clarity why it works. I won't bore you by going through the details of this. Becuase of our using advanced waves, we also had many apparent paradoxes, which we gradually reduced one by one, and saw that there was in fact no logical diffculty with the theory. It was perfectly satisfactory.

We also found that we could reformulate this thing in another way, and that is by a principle of least action. Since my original plan was to describe everything directly in terms of particle motions, it was my desire to represent this new theory without saying anything about fields. It turned out that we found a form for an action directly involving the motions of the charges only, which upon variation would give the equations of motion of these charges. The expression for this action A is

$$A=\Sigma_i m_i\int\left(\dot{X}^i{}_\mu\dot{X}^i{}_\mu\right)^{\frac{1}{2}}\mathrm{d}\alpha_i+\tfrac{1}{2}\sum_{\substack{ij\\ i\neq j}}e_ie_j\int\int\delta(I_{ij}{}^2)\,\dot{X}^i{}_\mu(\alpha_i)\,\dot{X}^j{}_\mu(\alpha_j)\,\mathrm{d}\alpha_i\,\mathrm{d}\alpha_j \quad (1)$$

where

$$I_{ij}{}^2=[X^i{}_\mu(\alpha_i)-X^j{}_\mu(\alpha_j)]\,[X^i{}_\mu(\alpha_i)-X^j{}_\mu(\alpha_j)]$$

where $X^i{}_\mu(\alpha_i)$ is the four-vector position of the i^{th} particle as a function of

some parameter α_i, $\dot{X}^i{}_\mu(\alpha_i)$ is $dX^i{}_\mu(\alpha)/d\alpha_i$. The first term is the integral of proper time, the ordinary action of relativistic mechanics of free particles of mass m_i. (We sum in the usual way on the repeated index μ.) The second term represents the electrical interaction of the charges. It is summed over each pair of charges (the factor $1/2$ is to count each pair once, the term $i=j$ is omitted to avoid self-action). The interaction is a double integral over a delta function of the square of space-time interval I^2 between two points on the paths. Thus, interaction occurs only when this interval vanishes, that is, along light cones.

The fact that the interaction is exactly one-half advanced and half-retarded meant that we could write such a principle of least action, whereas interaction *via* retarded waves alone cannot be written in such a way.

So, all of classical electrodynamics was contained in this very simple form. It looked good, and therefore, it was undoubtedly true, at least to the beginner. It automatically gave half-advanced and half-retarded effects and it was without fields. By omitting the term in the sum when $i=j$, *I* omit self-interaction and no longer have any infinite self-energy. This then was the hoped-for solution to the problem of ridding classical electrodynamics of the infinities.

It turns out, of course, that you can reinstate fields if you wish to, but you have to keep track of the field produced by each particle separately. This is because to find the right field to act on a given particle, you must exclude the field that it creates itself. A single universal field to which all contribute will not do. This idea had been suggested earlier by Frenkel and so we called these Frenkel fields. This theory which allowed only particles to act on each other was equivalent to Frenkel's fields using half-advanced and half-retarded solutions.

There were several suggestions for interesting modifications of electrodynamics. We discussed lots of them, but I shall report on only one. It was to replace this delta function in the interaction by another function, say, $f(I^2{}_{ij})$, which is not infinitely sharp. Instead of having the action occur only when the interval between the two charges is exactly zero, we would replace the delta function of I^2 by a narrow peaked thing. Let's say that $f(Z)$ is large only near $Z=0$ width of order a^2. Interactions will now occur when T^2-R^2 is of order a^2 roughly where T is the time difference and R is the separation of the charges. This might look like it disagrees with experience, but if a is some small distance, like 10^{-13} cm, it says that the time delay T in action is roughly $\sqrt{R^2 \pm a^2}$ or approximately,–if R is much larger than a, $T=R\pm a^2/2R$. This means that the deviation of time T from the ideal theoretical time R of Maxwell, gets smaller and smaller, the further the pieces are apart. Therefore, all theories

involving in analyzing generators, motors, etc., in fact, all of the tests of electrodynamics that were available in Maxwell's time, would be adequately satisfied if a were 10^{-13} cm. If R is of the order of a centimeter this deviation in T is only 10^{-26} parts. So, it was possible, also, to change the theory in a simple manner and to still agree with all observations of classical electrodynamics. You have no clue of precisely what function to put in for f, but it was an interesting possibility to keep in mind when developing quantum electrodynamics.

It also occurred to us that if we did that (replace δ by f) we could not reinstate the term $i=j$ in the sum because this would now represent in a relativistically invariant fashion a finite action of a charge on itself. In fact, it was possible to prove that if we did do such a thing, the main effect of the self-action (for not too rapid accelerations) would be to produce a modification of the mass. In fact, there need be no mass m_i, term, all the mechanical mass could be electromagnetic self-action. So, if you would like, we could also have another theory with a still simpler expression for the action A. In expression (1) only the second term is kept, the sum extended over all i and j, and some function f replaces δ. Such a simple form could represent all of classical electrodynamics, which aside from gravitation is essentially all of classical physics.

Although it may sound confusing, I am describing several different alternative theories at once. The important thing to note is that at this time we had all these in mind as different possibilities. There were several possible solutions of the difficulty of classical electrodynamics, any one of which might serve as a good starting point to the solution of the difficulties of quantum electrodynamics.

I would also like to emphasize that by this time I was becoming used to a physical point of view different from the more customary point of view. In the customary view, things are discussed as a function of time in very great detail. For example, you have the field at this moment, a differential equation gives you the field at the next moment and so on; a method, which I shall call the Hamilton method, the time differential method. We have, instead (in (1) say) a thing that describes the character of the path throughout all of space and time. The behavior of nature is determined by saying her whole space-time path has a certain character. For an action like (1) the equations obtained by variation (of $X^i{}_\mu(\alpha_i)$) are no longer at all easy to get back into Hamiltonian form. If you wish to use as variables only the coordinates of particles, then you can talk about the property of the paths–but the path of one particle at a given time is affected by the path of another at a different time. If you try to

masers is spotty. Presumably further work will allow interesting explorations in this region and a very fruitful, high-resolution spectroscopy.

Another region where coherent oscillators have not yet been developed is that of still shorter wavelengths stretching indefinitely beyond the near-ultraviolet, where the first such oscillators are now available. It can be shown that a rather severe and fundamental limitation exists as one proceeds to shorter wavelengths because of the continually increasing number of electromagnetic modes in a given volume and the faster and faster dissipation of energy into them by spontaneous emission.

Consider a cavity resonator of fixed volume, fixed-wall reflectivity, and fixed-fractional frequency-width $\Delta\nu/\nu$. Meeting the threshold condition (Eqn.11) in such a resonator requires that there is power which increases as ν^4, radiated by spontaneous emission into all modes of the system[30]. In the optical region this dissipated power for typical conditions, whereas at 50 ångstroms, in the soft X-ray region, it would be about 10^5 watts. The threshold condition would then be very difficult to maintain. But, by the same token, if it is maintained, the coherent X-ray beam produced would contain many kilowatts of power. It seems reasonable to expect, on this basis, that masers will be developed to wavelengths somewhat below 1 000 ångstroms, but that maser oscillations in the X-ray region will be very much more difficult.

Secondly, let us examine the monochromaticity which has been achieved. For the ammonia-beam maser, the variation of microwave oscillations was shown experimentally to agree with the theoretical expression (Eqn. 7) within the experimental precision of about 50 percent. This was done by beating two independent ammonia oscillators together and examining their relative phase variations[46]. A similar procedure can be carried out for two optical oscillators by mixing their two light beams together in a photocell and detecting the beat frequency. However, the technical difficulties in obtaining theoretical performance are rather more demanding than in the case of the ammonia maser. Equation 8 for a typical helium–neon laser predicts a frequency spread of about 10^{-2} cycle per second, or a fraction $3\cdot10^{-17}$ of the oscillation frequency of $3\cdot10^{14}$ cycles per second.

Almost all masers so far oscillating in the optical or near-infrared region require a sharper resonance, or higher Q, of the cavity than of the atomic resonance. Hence the frequency of oscillation is primarily determined, from Eqn. 5 by the cavity resonance. The frequency of oscillation thus depends on the separation L between mirrors, since from Eqn.12 $\nu = qc/2L$, where q is

describe, therefore, things differentially, telling what the present conditions of the particles are, and how these present conditions will affect the future– you see, it is impossible with particles alone, because something the particle did in the past is going to affect the future.

Therefore, you need a lot of bookkeeping variables to keep track of what the particle did in the past. These are called field variables. You will, also, have to tell what the field is at this present moment, if you are to be able to see later what is going to happen. From the overall space-time view of the least action principle, the field disappears as nothing but bookkeeping variables insisted on by the Hamiltonian method.

As a by-product of this same view, I received a telephone call one day at the graduate college at Princeton from Professor Wheeler, in which he said, «Feynman, I know why all electrons have the same charge and the same mass» «Why?» «Because, they are all the same electron!» And, then he explained on the telephone, «suppose that the world lines which we were ordinarily considering before in time and space–instead of only going up in time were a tremendous knot, and then, when we cut through the knot, by the plane corresponding to a fixed time, we would see many, many world lines and that would represent many electrons, except for one thing. If in one section this is an ordinary electron world line, in the section in which it reversed itself and is coming back from the future we have the wrong sign to the proper time – to the proper four velocities – and that's equivalent to changing the sign of the charge, and, therefore, that part of a path would act like a positron.» «But, Professor», I said, «there aren't as many positrons as electrons.» «Well, maybe they are hidden in the protons or something», he said. I did not take the idea that all the electrons were the same one from him as seriously as I took the observation that positrons could simply be represented as electrons going from the future to the past in a back section of their world lines. That, I stole!

To summarize, when I was done with this, as a physicist I had gained two things. One, I knew many different ways of formulating classical electrodynamics, with many different mathematical forms. I got to know how to express the subject every which way. Second, I had a point of view–the overall space-time point of view–and a disrespect for the Hamiltonian method of describing physics.

I would like to interrupt here to make a remark. The fact that electrodynamics can be written in so many ways–the differential equations of Maxwell, various minimum principles with fields, minimum principles without fields,

all different kinds of ways, was something I knew, but I have never understood. It always seems odd to me that the fundamental laws of physics, when discovered, can appear in so many different forms that are not apparently identical at first, but, with a little mathematical fiddling you can show the relationship. An example of that is the Schrödinger equation and the Heisenberg formulation of quantum mechanics. I don't know why this is – it remains a mystery, but it was something I learned from experience. There is always another way to say the same thing that doesn't look at all like the way you said it before. I don't know what the reason for this is. I think it is somehow a representation of the simplicity of nature. A thing like the inverse square law is just right to be represented by the solution of Poisson's equation, which, therefore, is a very different way to say the same thing that doesn't look at all like the way you said it before. I don't know what it means, that nature chooses these curious forms, but maybe that is a way of defining simplicity. Perhaps a thing is simple if you can describe it fully in several different ways without immediately knowing that you are describing the same thing.

I was now convinced that since we had solved the problem of classical electrodynamics (and completely in accordance with my program from M. I.T., only direct interaction between particles, in a way that made fields unnecessary) that everything was definitely going to be all right. I was convinced that all I had to do was make a quantum theory analogous to the classical one and everything would be solved.

So, the problem is only to make a quantum theory, which has as its classical analog, this expression (1). Now, there is no unique way to make a quantum theory from classical mechanics, although all the textbooks make believe there is. What they would tell you to do, was find the momentum variables and replace them by $(\hbar/i)(\partial/\partial x)$, but I couldn't find a momentum variable, as there wasn't any.

The character of quantum mechanics of the day was to write things in the famous Hamiltonian way – in the form of a differential equation, which described how the wave function changes from instant to instant, and in terms of an operator, H. If the classical physics could be reduced to a Hamiltonian form, everything was all right. Now, least action does not imply a Hamiltonian form if the action is a function of anything more than positions and velocities at the same moment. If the action is of the form of the integral of a function, (usually called the Lagrangian) of the velocities and positions at the same time

$$S = \int L(\dot{x}, x)\, dt \qquad (2)$$

then you can start with the Lagrangian and then create a Hamiltonian and work out the quantum mechanics, more or less uniquely. But this thing (1) involves the key variables, positions, at two different times and therefore, it was not obvious what to do to make the quantum-mechanical analogue.

I tried – I would struggle in various ways. One of them was this; if I had harmonic oscillators interacting with a delay in time, I could work out what the normal modes were and guess that the quantum theory of the normal modes was the same as for simple oscillators and kind of work my way back in terms of the original variables. I succeeded in doing that, but I hoped then to generalize to other than a harmonic oscillator, but I learned to my regret something, which many people have learned. The harmonic oscillator is too simple; very often you can work out what it should do in quantum theory without getting much of a clue as to how to generalize your results to other systems.

So that didn't help me very much, but when I was struggling with this problem, I went to a beer party in the Nassau Tavern in Princeton. There was a gentleman, newly arrived from Europe (Herbert Jehle) who came and sat next to me. Europeans are much more serious than we are in America because they think that a good place to discuss intellectual matters is a beer party. So, he sat by me and asked, «what are you doing» and so on, and I said, «I'm drinking beer.» Then I realized that he wanted to know what work I was doing and I told him I was struggling with this problem, and I simply turned to him and said, «listen, do you know any way of doing quantum mechanics, starting with action – where the action integral comes into the quantum mechanics?» «No», he said, «but Dirac has a paper in which the Lagrangian, at least, comes into quantum mechanics. I will show it to you tomorrow.»

Next day we went to the Princeton Library, they have little rooms on the side to discuss things, and he showed me this paper. What Dirac said was the following: There is in quantum mechanics a very important quantity which carries the wave function from one time to another, besides the differential equation but equivalent to it, a kind of a kernal, which we might call $K(x', x)$, which carries the wave function $\psi(x)$ known at time t, to the wave function $\psi(x')$ at time, $t+\varepsilon$. Dirac points out that this function K was *analogous* to the quantity in classical mechanics that you would calculate if you took the exponential of $i\varepsilon$, multiplied by the Lagrangian $L(\dot{x}, x)$ imagining that these two positions x,x' corresponded t and $t+\varepsilon$. In other words,

$$K(x', x) \text{ is analogous to } e^{i\varepsilon L(\frac{x'-x}{\varepsilon}, x)/\hbar}$$

Professor Jehle showed me this, I read it, he explained it to me, and I said, «what does he mean, they are analogous; what does that mean, *analogous*? What is the use of that?» He said, «you Americans! You always want to find a use for everything!» I said, that I thought that Dirac must mean that they were equal. «No», he explained, «he doesn't mean they are equal.» «Well», I said, «let's see what happens if we make them equal.»

So I simply put them equal, taking the simplest example where the Lagrangian is $^1/_2 Mx^2 - V(x)$ but soon found I had to put a constant of proportionality A in, suitably adjusted. When I substituted $Ae^{i\varepsilon L/\hbar}$ for K to get

$$\psi(x', t+\varepsilon) = \int A \exp\left[\frac{i\varepsilon}{\hbar} L\left(\frac{x'-x}{\varepsilon}, x\right)\right] \psi(x, t)\, dx \qquad (3)$$

and just calculated things out by Taylor series expansion, out came the Schrödinger equation. So, I turned to Professor Jehle, not really understanding, and said, «well, you see Professor Dirac meant that they were proportional.» Professor Jehle's eyes were bugging out – he had taken out a little notebook and was rapidly copying it down from the blackboard, and said, «no, no, this is an important discovery. You Americans are always trying to find out how something can be used. That's a good way to discover things!» So, I thought I was finding out what Dirac meant, but, as a matter of fact, had made the discovery that what Dirac thought was analogous, was, in fact, equal. I had then, at least, the connection between the Lagrangian and quantum mechanics, but still with wave functions and infinitesimal times.

It must have been a day or so later when I was lying in bed thinking about these things, that I imagined what would happen if I wanted to calculate the wave function at a finite interval later.

I would put one of these factors $e^{i\varepsilon L}$ in here, and that would give me the wave functions the next moment, $t+\varepsilon$ and then I could substitute that back into (3) to get another factor of $e^{i\varepsilon L}$ and give me the wave function the next moment, $t+2\varepsilon$, and so on and so on. In that way I found myself thinking of a large number of integrals, one after the other in sequence. In the integrand was the product of the exponentials, which, of course, was the exponential of the sum of terms like εL. Now, L is the Lagrangian and ε is like the time interval dt, so that if you took a sum of such terms, that's exactly like an integral. That's like Riemann's formula for the integral $\int L dt$, you just take the value at each point and add them together. We are to take the limit as $\varepsilon - 0$, of course. Therefore, the connection between the wave function of one instant and the wave function of another instant a finite time later could be obtained by an infinite number of integrals, (because ε goes to zero, of course) of ex-

ponential $(iS/\hbar)$ where S is the action expression (2). At last, I had succeeded in representing quantum mechanics directly in terms of the action S.

This led later on to the idea of the amplitude for a path; that for each possible way that the particle can go from one point to another in space-time, there's an amplitude. That amplitude is e to the $i/\hbar$ times the action for the path. Amplitudes from various paths superpose by addition. This then is another, a third way, of describing quantum mechanics, which looks quite different than that of Schrödinger or Heisenberg, but which is equivalent to them.

Now immediately after making a few checks on this thing, what I wanted to do, of course, was to substitute the action (1) for the other (2). The first trouble was that I could not get the thing to work with the relativistic case of spin one-half. However, although I could deal with the matter only non-relativistically, I could deal with the light or the photon interactions perfectly well by just putting the interaction terms of (1) into any action, replacing the mass terms by the non-relativistic $(M\dot{x}^2/2)dt$. When the action has a delay, as it now had, and involved more than one time, I had to lose the idea of a wave function. That is, I could no longer describe the program as; given the amplitude for all positions at a certain time to compute the amplitude at another time. However, that didn't cause very much trouble. It just meant developing a new idea. Instead of wave functions we could talk about this; that if a source of a certain kind emits a particle, and a detector is there to receive it, we can give the amplitude that the source will emit and the detector receive. We do this without specifying the exact instant that the source emits or the exact instant that any detector receives, without trying to specify the state of anything at any particular time in between, but by just finding the amplitude for the complete experiment. And, then we could discuss how that amplitude would change if you had a scattering sample in between, as you rotated and changed angles, and so on, without really having any wave functions.

It was also possible to discover what the old concepts of energy and momentum would mean with this generalized action. And, so I believed that I had a quantum theory of classical electrodynamics–or rather of this new classical electrodynamics described by action (1). I made a number of checks. If I took the Frenkel field point of view, which you remember was more differential, I could convert it directly to quantum mechanics in a more conventional way. The only problem was how to specify in quantum mechanics the classical boundary conditions to use only half-advanced and half-retarded solutions. By some ingenuity in defining what that meant, I found that the quantum

mechanics with Frenkel fields, plus a special boundary condition, gave me back this action, (1) in the new form of quantum mechanics with a delay. So, various things indicated that there wasn't any doubt I had everything straightened out.

It was also easy to guess how to modify the electrodynamics, if anybody ever wanted to modify it. I just changed the delta to an f, just as I would for the classical case. So, it was very easy, a simple thing. To describe the old retarded theory without explicit mention of fields I would have to write probabilities, not just amplitudes. I would have to square my amplitudes and that would involve double path integrals in which there are two S's and so forth. Yet, as I worked out many of these things and studied different forms and different boundary conditions. I got a kind of funny feeling that things weren't exactly right. I could not clearly identify the difficulty and in one of the short periods during which I imagined I had laid it to rest, I published a thesis and received my Ph. D.

During the war, I didn't have time to work on these things very extensively, but wandered about on buses and so forth, with little pieces of paper, and struggled to work on it and discovered indeed that there was something wrong, something terribly wrong. I found that if one generalized the action from the nice Langrangian forms (2) to these forms (1) then the quantities which I defined as energy, and so on, would be complex. The energy values of stationary states wouldn't be real and probabilities of events wouldn't add up to 100%. That is, if you took the probability that this would happen and that would happen –everything you could think of would happen, it would not add up to one.

Another problem on which I struggled very hard, was to represent relativistic electrons with this new quantum mechanics. I wanted to do a unique and different way–and not just by copying the operators of Dirac into some kind of an expression and using some kind of Dirac algebra instead of ordinary complex numbers. I was very much encouraged by the fact that in one space dimension, I did find a way of giving an amplitude to every path by limiting myself to paths, which only went back and forth at the speed of light. The amplitude was simple ($i\varepsilon$) to a power equal to the number of velocity reversals where I have divided the time into steps ε and I am allowed to reverse velocity only at such a time. This gives (as ε approaches zero) Dirac's equation in two dimensions–one dimension of space and one of time ($\hbar = M = c = 1$).

Dirac's wave function has four components in four dimensions, but in this case, it has only two components and this rule for the amplitude of a path

automatically generates the need for two components. Because if this is the formula for the amplitudes of path, it will not do you any good to know the total amplitude of all paths, which come into a given point to find the amplitude to reach the next point. This is because for the next time, if it came in from the right, there is no new factor $i\epsilon$ if it goes out to the right, whereas, if it came in from the left there was a new factor $i\epsilon$. So, to continue this same information forward to the next moment, it was not sufficient information to know the total amplitude to arrive, but you had to know the amplitude to arrive from the right and the amplitude to arrive to the left, independently. If you did, however, you could then compute both of those again independently and thus you had to carry two amplitudes to form a differential equation (first order in time).

And, so I dreamed that if I were clever, I would find a formula for the amplitude of a path that was beautiful and simple for three dimensions of space and one of time, which would be equivalent to the Dirac equation, and for which the four components, matrices, and all those other mathematical funny things would come out as a simple consequence–I have never succeeded in that either. But, I did want to mention some of the unsuccessful things on which I spent almost as much effort, as on the things that did work.

To summarize the situation a few years after the way, I would say, I had much experience with quantum electrodynamics, at least in the knowledge of many different ways of formulating it, in terms of path integrals of actions and in other forms. One of the important by-products, for example, of much experience in these simple forms, was that it was easy to see how to combine together what was in those days called the longitudinal and transverse fields, and in general, to see clearly the relativistic invariance of the theory. Because of the need to do things differentially there had been, in the standard quantum electrodynamics, a complete split of the field into two parts, one of which is called the longitudinal part and the other mediated by the photons, or transverse waves. The longitudinal part was described by a Coulomb potential acting instantaneously in the Schrödinger equation, while the transverse part had entirely different description in terms of quantization of the transverse waves. This separation depended upon the relativistic tilt of your axes in space-time. People moving at different velocities would separate the same field into longitudinal and transverse fields in a different way. Furthermore, the entire formulation of quantum mechanics insisting, as it did, on the wave function at a given time, was hard to analyze relativistically. Somebody else in a different coordinate system would calculate the succession of events in terms of wave

functions on differently cut slices of space-time, and with a different separation of longitudinal and transverse parts. The Hamiltonian theory did not look relativistically invariant, although, of course, it was. One of the great advantages of the overall point of view, was that you could see the relativistic invariance right away – or as Schwinger would say – the covariance was manifest. I had the advantage, therefore, of having a manifestedly covariant form for quantum electrodynamics with suggestions for modifications and so on. I had the disadvantage that if I took it too seriously – I mean, if I took it seriously at all in this form, – I got into trouble with these complex energies and the failure of adding probabilities to one and so on. I was unsuccessfully struggling with that.

Then Lamb did his experiment, measuring the separation of the $2S_{\frac{1}{2}}$ and $2P_{\frac{1}{2}}$ levels of hydrogen, finding it to be about 1000 megacycles of frequency difference. Professor Bethe, with whom I was then associated at Cornell, is a man who has this characteristic: If there's a good experimental number you've got to figure it out from theory. So, he forced the quantum electrodynamics of the day to give him an answer to the separation of these two levels. He pointed out that the self-energy of an electron itself is infinite, so that the calculated energy of a bound electron should also come out infinite. But, when you calculated the separation of the two energy levels in terms of the corrected mass instead of the old mass, it would turn out, he thought, that the theory would give convergent finite answers. He made an estimate of the splitting that way and found out that it was still divergent, but he guessed that was probably due to the fact that he used an unrelativistic theory of the matter. Assuming it would be convergent if relativistically treated, he estimated he would get about a thousand megacycles for the Lamb-shift, and thus, made the most important discovery in the history of the theory of quantum electrodynamics. He worked this out on the train from Ithaca, New York to Schenectady and telephoned me excitedly from Schenectady to tell me the result, which I don't remember fully appreciating at the time.

Returning to Cornell, he gave a lecture on the subject, which I attended. He explained that it gets very confusing to figure out exactly which infinite term corresponds to what in trying to make the correction for the infinite change in mass. If there were any modifications whatever, he said, even though not physically correct, (that is not necessarily the way nature actually works) but any modification whatever at high frequencies, which would make this correction finite, then there would be no problem at all to figuring out how to keep track of everything. You just calculate the finite mass correc-

tion Δm to the electron mass m_o, substitute the numerical values of $m_o + \Delta m$ for m in the results for any other problem and all these ambiguities would be resolved. If, in addition, this method were relativistically invariant, then we would be absolutely sure how to do it without destroying relativistically invariant.

After the lecture, I went up to him and told him, «I can do that for you, I'll bring it in for you tomorrow.» I guess I knew every way to modify quantum electrodynamics known to man, at the time. So, I went in next day, and explained what would correspond to the modification of the delta-function to f and asked him to explain to me how you calculate the self-energy of an electron, for instance, so we can figure out if it's finite.

I want you to see an interesting point. I did not take the advice of Professor Jehle to find out how it was useful. I never used all that machinery which I had cooked up to solve a single relativistic problem. I hadn't even calculated the self-energy of an electron up to that moment, and was studying the difficulties with the conservation of probability, and so on, without actually doing anything, except discussing the general properties of the theory.

But now I went to Professor Bethe, who explained to me on the blackboard, as we worked together, how to calculate the self-energy of an electron. Up to that time when you did the integrals they had been logarithmically divergent. I told him how to make the relativistically invariant modifications that I thought would make everything all right. We set up the integral which then diverged at the sixth power of the frequency instead of logarithmically!

So, I went back to my room and worried about this thing and went around in circles trying to figure out what was wrong because I was sure physically everything had to come out finite, I couldn't understand how it came out infinite. I became more and more interested and finally realized I had to learn how to make a calculation. So, ultimately, I taught myself how to calculate the self-energy of an electron working my patient way through the terrible confusion of those days of negative energy states and holes and longitudinal contributions and so on. When I finally found out how to do it and did it with the modifications I wanted to suggest, it turned out that it was nicely convergent and finite, just as I had expected. Professor Bethe and I have never been able to discover what we did wrong on that blackboard two months before, but apparently we just went off somewhere and we have never been able to figure out where. It turned out, that what I had proposed, if we had carried it out without making a mistake would have been all right and would have given a finite correction. Anyway, it forced me to go back over all this and to

convince myself physically that nothing can go wrong. At any rate, the correction to mass was now finite, proportional to $\ln(ma/\hbar)$ where a is the width of that function f which was substituted for δ. If you wanted an unmodified electrodynamics, you would have to take a equal to zero, getting an infinite mass correction. But, that wasn't the point. Keeping a finite, I simply followed the program outlined by Professor Bethe and showed how to calculate all the various things, the scatterings of electrons from atoms without radiation, the shifts of levels and so forth, calculating everything in terms of the experimental mass, and noting that the results as Bethe suggested, were not sensitive to a in this form and even had a definite limit as $a \to 0$.

The rest of my work was simply to improve the techniques then available for calculations, making diagrams to help analyze perturbation theory quicker. Most of this was first worked out by guessing–you see, I didn't have the relativistic theory of matter. For example, it seemed to me obvious that the velocities in non-relativistic formulas have to be replaced by Dirac's matrix α or in the more relativistic forms by the operators γ_μ. I just took my guesses from the forms that I had worked out using path integrals for non-relativistic matter, but relativistic light. It was easy to develop rules of what to substitute to get the relativistic case. I was very surprised to discover that it was not known at that time, that every one of the formulas that had been worked out so patiently by separating longitudinal and transverse waves could be obtained from the formula for the transverse waves alone, if instead of summing over only the two perpendicular polarization directions you would sum over all four possible directions of polarization. It was so obvious from the action (1) that I thought it was general knowledge and would do it all the time. I would get into arguments with people, because I didn't realize they didn't know that; but, it turned out that all their patient work with the longitudinal waves was always equivalent to just extending the sum on the two transverse directions of polarization over all four directions. This was one of the amusing advantages of the method. In addition, I included diagrams for the various terms of the perturbation series, improved notations to be used, worked out easy ways to evaluate integrals, which occurred in these problems, and so on, and made a kind of handbook on how to do quantum electrodynamics.

But one step of importance that was physically new was involved with the negative energy sea of Dirac, which caused me so much logical difficulty. I got so confused that I remembered Wheeler's old idea about the positron being, maybe, the electron going backward in time. Therefore, in the time depen-

dent perturbation theory that was usual for getting self-energy, I simply supposed that for a while we could go backward in the time, and looked at what terms I got by running the time variables backward. They were the same as the terms that other people got when they did the problem a more complicated way, using holes in the sea, except, possibly, for some signs. These, I, at first, determined empirically by inventing and trying some rules.

I have tried to explain that all the improvements of relativistic theory were at first more or less straightforward, semi-empirical shenanigans. Each time I would discover something, however, I would go back and I would check it so many ways, compare it to every problem that had been done previously in electrodynamics (and later, in weak coupling meson theory) to see if it would always agree, and so on, until I was absolutely convinced of the truth of the various rules and regulations which I concocted to simplify all the work.

During this time, people had been developing meson theory, a subject I had not studied in any detail. I became interested in the possible application of my methods to perturbation calculations in meson theory. But, what was meson theory? All I knew was that meson theory was something analogous to electrodynamics, except that particles corresponding to the photon had a mass. It was easy to guess the δ-function in (1), which was a solution of d'Alembertian equals zero, was to be changed to the corresponding solution of d'Alembertian equals m^2. Next, there were different kind of mesons–the one in closest analogy to photons, coupled *via* $\gamma_\mu\gamma_\mu$, are called vector mesons–there were also scalar mesons. Well, maybe that corresponds to putting unity in place of the γ_μ, I would here then speak of «pseudo vector coupling» and I would guess what that probably was. I didn't have the knowledge to understand the way these were defined in the conventional papers because they were expressed at that time in terms of creation and annihilation operators, and so on, which, I had not successfully learned. I remember that when someone had started to teach me about creation and annihilation operators, that this operator creates an electron, I said, «how do you create an electron? It disagrees with the conservation of charge», and in that way, I blocked my mind from learning a very practical scheme of calculation. Therefore, I had to find as many opportunities as possible to test whether I guessed right as to what the various theories were.

One day a dispute arose at a Physical Society meeting as to the correctness of a calculation by Slotnick of the interaction of an electron with a neutron using pseudo scalar theory with pseudo vector coupling and also, pseudo scalar theory with pseudo scalar coupling. He had found that the answers were not

the same, in fact, by one theory, the result was divergent, although convergent with the other. Some people believed that the two theories must give the same answer for the problem. This was a welcome opportunity to test my guesses as to whether I really did understand what these two couplings were. So, I went home, and during the evening I worked out the electron neutron scattering for the pseudo scalar and pseudo vector coupling, saw they were not equal and subtracted them, and worked out the difference in detail. The next day at the meeting, I saw Slotnick and said, «Slotnick, I worked it out last night, I wanted to see if I got the same answers you do. I got a different answer for each coupling–but, I would like to check in detail with you because I want to make sure of my methods.» And, he said, «what do you mean you worked it out last night, it took me six months!» And, when we compared the answers he looked at mine and he asked, «what is that Q in there, that variable Q?» (I had expressions like $(\tan^{-1}Q)/Q$ etc.). I said, «that's the momentum transferred by the electron, the electron deflected by different angles.» «Oh», he said, «no, I only have the limiting value as Q approaches zero; the forward scattering.» Well, it was easy enough to just substitute Q equals zero in my form and I then got the same answers as he did. But, it took him six months to do the case of zero momentum transfer, whereas, during one evening I had done the finite and arbitrary momentum transfer. That was a thrilling moment for me, like receiving the Nobel Prize, because that convinced me, at last, I did have some kind of method and technique and understood how to do something that other people did not know how to do. That was my moment of triumph in which I realized I really had succeeded in working out something worthwhile.

At this stage, I was urged to publish this because everybody said it looks like an easy way to make calculations, and wanted to know how to do it. I had to publish it, missing two things; one was proof of every statement in a mathematically conventional sense. Often, even in a physicist's sense, I did not have a demonstration of how to get all of these rules and equations from conventional electrodynamics. But, I did know from exprience, from fooling around, that everything was, in fact, equivalent to the regular electrodynamics and had partial proofs of many pieces, although, I never really sat down, like Euclid did for the geometers of Greece, and made sure that you could get it all from a single simple set of axioms. As a result, the work was criticized, I don't know whether favorably or unfavorably, and the «method» was called the «intuitive method». For those who do not realize it, however, I should like to emphasize that there is a lot of work involved in using this «intuitive

method» successfully. Because no simple clear proof of the formula or idea presents itself, it is necessary to do an unusually great amount of checking and rechecking for consistency and correctness in terms of what is known, by comparing to other analogous examples, limiting cases, etc. In the face of the lack of direct mathematical demonstration, one must be careful and thorough to make sure of the point, and one should make a perpetual attempt to demonstrate as much of the formula as possible. Nevertheless, a very great deal more truth can become known than can be proven.

It must be clearly understood that in all this work, I was representing the conventional electrodynamics with retarded interaction, and not my half-advanced and half-retarded theory corresponding to (1). I merely use (1) to guess at forms. And, one of the forms I guessed at corresponded to changing δ to a function f of width a^2, so that I could calculate finite results for all of the problems. This brings me to the second thing that was missing when I published the paper, an unresolved difficulty. With δ replaced by f the calculations would give results which were not «unitary», that is, for which the sum of the probabilities of all alternatives was not unity. The deviation from unity was very small, in practice, if a was very small. In the limit that I took a very tiny, it might not make any difference. And, so the process of the renormalization could be made, you could calculate everything in terms of the experimental mass and then take the limit and the apparent difficulty that the unitary is violated temporarily seems to disappear. I was unable to demonstrate that, as a matter of fact, it does.

It is lucky that I did not wait to straighten out that point, for as far as I know, nobody has yet been able to resolve this question. Experience with meson theories with stronger couplings and with strongly coupled vector photons, although not proving anything, convinces me that if the coupling were stronger, or if you went to a higher order (137th order of perturbation theory for electrodynamics), this difficulty would remain in the limit and there would be real trouble. That is, I believe there is really no satisfactory quantum electrodynamics, but I'm not sure. And, I believe, that one of the reasons for the slowness of present-day progress in understanding the strong interactions is that there isn't any relativistic theoretical model, from which you can really calculate everything. Although, it is usually said, that the difficulty lies in the fact that *strong* interactions are too hard to calculate, I believe, it is really because strong interactions in field theory have no solution, have no sense–they're either infinite, or, if you try to modify them, the modification destroys the unitarity. I don't think we have a completely satisfactory relativistic quan-

tum-mechanical model, even one that doesn't agree with nature, but, at least, agrees with the logic that the sum of probability of all alternatives has to be 100%. Therefore, I think that the renormalization theory is simply a way to sweep the difficulties of the divergences of electrodynamics under the rug. I am, of course, not sure of that.

This completes the story of the development of the space-time view of quantum electrodynamics. I wonder if anything can be learned from it. I doubt it. It is most striking that most of the ideas developed in the course of this research were not ultimately used in the final result. For example, the half-advanced and half-retarded potential was not finally used, the action expression (1) was not used, the idea that charges do not act on themselves was abandoned. The path-integral formulation of quantum mechanics was useful for guessing at final expressions and at formulating the general theory of electrodynamics in new ways–although, strictly it was not absolutely necessary. The same goes for the idea of the positron being a backward moving electron, it was very convenient, but not strictly necessary for the theory because it is exactly equivalent to the negative energy sea point of view.

We are struck by the very large number of different physical viewpoints and widely different mathematical formulations that are all equivalent to one another. The method used here, of reasoning in physical terms, therefore, appears to be extremely inefficient. On looking back over the work, I can only feel a kind of regret for the enormous amount of physical reasoning and mathematically re-expression which ends by merely re-expressing what was previously known, although in a form which is much more efficient for the calculation of specific problems. Would it not have been much easier to simply work entirely in the mathematical framework to elaborate a more efficient expression? This would certainly seem to be the case, but it must be remarked that although the problem actually solved was only such a reformulation, the problem originally tackled was the (possibly still unsolved) problem of avoidance of the inifinities of the usual theory. Therefore, a new theory was sought, not just a modification of the old. Although the quest was unsuccessful, we should look at the question of the value of physical ideas in developing a *new* theory.

Many different physical ideas can describe the same physical reality. Thus, classical electrodynamics can be described by a field view, or an action at a distance view, etc. Originally, Maxwell filled space with idler wheels, and Faraday with fields lines, but somehow the Maxwell equations themselves are

pristine and independent of the elaboration of words attempting a physical description. The only true physical description is that describing the experimental meaning of the quantities in the equation–or better, the way the equations are to be used in describing experimental observations. This being the case perhaps the best way to proceed is to try to guess equations, and disregard physical models or descriptions. For example, McCullough guessed the correct equations for light propagation in a crystal long before his colleagues using elastic models could make head or tail of the phenomena, or again, Dirac obtained his equation for the description of the electron by an almost purely mathematical proposition. A simple physical view by which all the contents of this equation can be seen is still lacking.

Therefore, I think equation guessing might be the best method to proceed to obtain the laws for the part of physics which is presently unknown. Yet, when I was much younger, I tried this equation guessing and I have seen many students try this, but it is very easy to go off in wildly incorrect and impossible directions. I think the problem is not to find the *best* or most efficient method to proceed to a discovery, but to find any method at all. Physical reasoning does help some people to generate suggestions as to how the unknown may be related to the known. Theories of the known, which are described by different physical ideas may be equivalent in all their predictions and are hence scientifically indistinguishable. However, they are not psychologically identical when trying to move from that base into the unknown. For different views suggest different kinds of modifications which might be made and hence are not equivalent in the hypotheses one generates from them in ones attempt to understand what is not yet understood. I, therefore, think that a good theoretical physicist today might find it useful to have a wide range of physical viewpoints and mathematical expressions of the same theory (for example, of quantum electrodynamics) available to him. This may be asking too much of one man. Then new students should as a class have this. If every individual student follows the same current fashion in expressing and thinking about electrodynamics or field theory, then the variety of hypotheses being generated to understand strong interactions, say, is limited. Perhaps rightly so, for possibly the chance is high that the truth lies in the fashionable direction. But, on the off-chance that it is in another direction–a direction obvious from an unfashionable view of field theory–who will find it? Only someone who has sacrificed himself by teaching himself quantum electrodynamics from a peculiar and unusual point of view; one that he may have to invent for himself. I say sacrificed himself because he most likely will get

nothing from it, because the truth may lie in another direction, perhaps even the fashionable one.

But, if my own experience is any guide, the sacrifice is really not great because if the peculiar viewpoint taken is truly experimentally equivalent to the usual in the realm of the known there is always a range of applications and problems in this realm for which the special viewpoint gives one a special power and clarity of thought, which is valuable in itself. Furthermore, in the search for new laws, you always have the psychological excitement of feeling that possible nobody has yet thought of the crazy possibility you are looking at right now.

So what happened to the old theory that I fell in love with as a youth? Well, I would say it's become an old lady, that has vey little attractive left in her and the young today will not have their hearts pound when they look at her anymore. But, we can say the best we can for any old woman, that she has been a very good mother and she has given birth to some very good children. And, I thank the Swedish Academy of Sciences for complimenting one of them. Thank you.

Biography

Richard P. Feynman was born in New York City on the 11th May 1918. He studied at the Massachusetts Institute of Technology where he obtained his B. Sc. in 1939 and at Princeton University where he obtained his Ph. D. in 1942. He was Research Assistant at Princeton (1940–1941), Professor of Theoretical Physics at Cornell University (1945–1950), Visiting Professor and thereafter appointed Professor of Theoretical Physics at the California Institute of Technology (1950–1959). At present he is Richard Chace Tolman Professor of Theoretical Physics at the California Institute of Technology.

Professor Feynman is a member of the American Physical Society, the American Association for the Advancement of Science; the National Academy of Science; in 1965 he was elected a foreign member of the Royal Society, London (Great Britain).

He holds the following awards: Albert Einstein Award (1954, Princeton); Einstein Award (Albert Einstein Award College of Medicine); Lawrence Award (1962).

Richard Feynman is married to Gweneth Howarth, they have a son, Carl Richard (born 22nd April 1961), and a daughter Michelle Catherine (born 13th August 1968).

Physics 1966

181

ALFRED KASTLER

«for the discovery and development of optical methods for studying Hertzian resonances in atoms»

Physics 1966

Presentation Speech by Professor Ivar Waller, member of the Nobel Committee for Physics

Your Majesty, Royal Highnesses, Ladies and Gentlemen.

When, shortly after 1930, Alfred Kastler embarked upon a scientific career, he concentrated his attention on problems connected with light scattering. He used novel methods to analyse this phenomenon, which had already been studied by projecting light emitted by certain atoms into a chamber containing the same kind of atoms. The illuminated atoms are thus excited by the light to a higher energy level. When a resonance effect of this kind is produced, strong fluorescence is emitted by the excited atoms as they return to the ground state.

The phenomenon received close attention a little earlier, particularly after it was found that the fluorescence is strongly polarized by placing a polarizer between the lamp and the resonance chamber. Another observation was that this polarization was considerably influenced by a magnetic field acting on the illuminated atoms.

Kastler made an important contribution to our understanding of these phenomena. He studied the relationships between the spatial orientation of the atoms and the polarization of their radiation, and thus laid the foundations of the work that is today honoured with the Nobel Physics prize.

The starting point of the work was research into Hertzian resonances. These are produced when atoms interact with radio waves or microwaves, *i.e.* with electromagnetic radiation having a frequency at least a thousand times lower than visible light. Such waves are therefore well suited to the study of fine details in spectra, which, though observable by optical spectroscopy, could not be measured with satisfactory precision by this method. Hertzian resonances were first used for this purpose – and with success – in 1938, by Rabi following Gorter's suggestion. Rabi was able to measure, with high precision, the splitting of energy levels into a number of sublevels, a phenomenon that is produced in the presence of a magnetic field and that is due to the orientation of the atoms in space. The hyperfine structure is another kind of small subdivisions, associated with the magnetic and electric moments of magnetic nuclei. On the basis of his exact measurements, Rabi was in a position to calculate these nuclear moments with great precision.

Aided by Jean Brossel, first his pupil and later close coworker, Kastler was the first to propose a method of investigating Hertzian resonances by optical methods, indicating the possibility of exciting selectively magnetic sublevels from excited states by polarized light having the resonance frequency. If a high-frequency oscillating magnetic field is applied, Hertzian resonance will be induced when the ratio of this frequency to an applied constant magnetic field is suitably chosen. Hertzian resonances tend to equalize the population of the magnetic sublevels, and inconsequence influence the observed polarization of the fluorescence. In practice, the resonance chamber in the process described earlier is surrounded by a coil carrying a current of radio or microwave frequency.

The experiment was carried out some years later by Brossel in collaboration with the American physicist Bitter. To extend the use of Hertzian resonances to excited states Bitter had already suggested combining optical and Hertzian resonances, but he did not propose a method of accomplishing this aim. He called the Brossel–Kastler method double optical resonance.

New profound analysis of the atomic processes connected with the scattering of resonance radiation led Kastler to the method of optical pumping, which he proposed in 1950. In this method the atoms are illuminated with resonance radiation, which is as a rule circularly polarized. According to Kastler, the atoms returning to the ground state concentrate in certain sublevels and assume preferential orientations in space if the experiment is conducted under appropriate conditions. The use of this method should allow orientation of both atoms and atomic nuclei. The experiment was actually performed two years later by Brossel, Kastler and Winter.

Double resonance and optical pumping permit very sensitive detection of Hertzian resonances, because such resonances provoke easily observable optical effects. These methods are therefore based on a different principle than ESR or NMR spectroscopy; in contrast to the latter methods, they can be applied to materials having very low density. The methods were systematically developed by Kastler in collaboration with Brossel and with a large number of young and brilliant researchers, and the investigations bear witness to the extraordinary fertility and the numerous possibilities of application of this approach.

As an important example of the phenomena involving excited states studied by double resonance in Kastler's laboratory, I shall mention the narrowing of spectral lines with increasing gas pressure within the resonance chamber.

Experiments on optical pumping were at first done with atomic beams.

They led to extensive experimental and theoretical investigations of the simultaneous interactions of several quanta of an oscillating magnetic field with atoms. An important improvement in the method of pumping was obtained when the attempts to conduct these experiments on the vapour in the resonance chamber proved successful. Some very interesting work was done on the relaxation of atoms back to the disordered state after pumping, which provided information on the mechanism acting in interatomic collisions and in collisions between atoms and the walls of the container.

In the last few years, Cohen-Tannoudji has conducted research of extreme general importance, again in Kastler's laboratory, by studying the broadening and displacement of energy levels in pumped atoms, caused by their interactions with an electromagnetic field.

A large number of nuclear moments have been determined with high precision. Kastler's ideas about optical pumping played an important part in the development of the laser. Optical pumping has permitted the construction of easy to use and very sensitive magnetometers as well as atomic clocks.

Professor Alfred Kastler. Through your discoveries, made partly in collaboration with your erstwhile pupil Jean Brossel, you have set a seal upon the great French tradition in optical science. Your methods have been perfected and have been successfully applied to a large number of fundamental problems by yourself and by the team of eminent young scientists attracted by the illustrious reputation of your laboratory. You have consistently acknowledged the research of your colleagues with characteristic generosity and personal modesty.

I ask you, Professor Kastler, to receive the Nobel Prize for Physics from the hands of His Majesty the King.

ALFRED KASTLER

Optical methods for studying Hertzian resonances

Nobel Lecture, December 12, 1966

During my first year of studies at the École Normale Supérieure in Paris, our teacher, Eugène Bloch, introduced us to quantum physics, which at that time was little taught in France. Like he, I was of Alsatian extraction and knew German. He strongly advised me to read Sommerfeld's admirable book *Atombau und Spektrallinien*[1]. In the course of this reading, I became particularly interested in the application of the principle of conservation of angular momentum during interactions between electromagnetic radiation and atoms, an application which had led A. Rubinowicz[2] to the interpretation of the selection rules for the azimuthal quantum number and polarization in the Zeeman effect. In the hypothesis of light quanta, this principle attributed to the photons a momentum $+\hbar$ or $-\hbar$ according to whether the light was polarized circularly to the right (σ^+) or to the left (σ^-), natural light being a mixture of the two kinds of photons.

In 1931, W. Hanle and R. Bär[3] independently discovered an interesting characteristic of Raman spectra. The study of the polarization of Raman lines at right angles to the incident beam made it possible to classify the Raman lines of a molecule into two categories: «depolarized» lines with a depolarization factor of 6/7 and «polarized» lines, whose polarization was generally appreciable. Placzek's theory had attributed the former to periodic molecular motions which modify the symmetry elements the molecule possesses at rest, among which are included rotational Raman lines, and the latter to totally symmetric vibrations which maintain the symmetry elements of the molecule at rest.

Hanle and Bär illuminated the medium with circularly polarized incident light and observed that, under these conditions, the Raman lines scattered longitudinally had the same circular polarization as the incident light in the case of totally symmetric vibrations, but that the direction of circular polarization was reversed for lines not totally symmetrical. In a note[4], I pointed out that for rotational lines this curious result was an immediate consequence of the principle of conservation of angular momentum applied to light scattering.

At about the same time, Jean Cabannes[5] explained the Hanle and Bär result by the classical polarizability theory, but these publications had been preceded

by an article of Raman and Bhagavantam[6] who saw proof of the existence of photon spin in the experimental results cited.

At the time, another experiment seemed to me appropriate for demonstrating the possible existence of a transverse component of the momentum of photons: the study of linearly polarized light originating from a rotating atomic oscillator and viewed edge on. This case arises for the σ components of the transverse Zeeman effect, which correspond to the σ^+ and σ^- components of the longitudinal effect. The experiment[7] that I performed during the Easter vacation of 1931 at the Physics Laboratory of the École Normale Supérieure in Paris, with the aid of Felix Esclangon, was a failure: there is no transverse component of angular momentum in light. Here again, I had been preceded by R. Frisch[8], who had reached similar conclusions.

These initial attempts caused me to examine more systematically the consequences of the principle of conservation of angular momentum in light scattering and in fluorescence[9]. I realized that the optical excitation of atoms in steps[10,11] constituted a particularly interesting field of application since, in this case, the operator is free to polarize the different monochromatic radiations whose absorption raises the atom through the successive steps of increasing energy. My thesis consisted in applying this method to the mercury atom[12]. It enabled me to check out the various predictions. It constituted a first attempt to obtain, by suitable polarization of the exciting radiation, a selective excitation of definite magnetic sublevels. The very fact that the fluorescence intensity resulting from a step excitation is of nonnegligible order of magnitude relative to the emission intensity resulting from a single excitation showed me, in addition, that the population obtained in the course of a stationary irradiation in the first excited state may become a nonnegligible fraction of the population of the ground state despite the weak intensity of the monochromatic light sources available at that time.

After the development of methods of Hertzian resonance of the ground state of isolated atoms by I. Rabi and his students[13] and after the first and famous application by Lamb and Retherford[14] of these methods to the states $n=2$ of the hydrogen atom, the American physicist Francis Bitter attracted attention to the interest inherent in extending the techniques of radio-frequency spectroscopy to the excited states of atoms; but the method he proposed for doing this[15] proved to be inexact[16]. My former student Jean Brossel was then working under the direction of Bitter at M. I. T. After an exchange of correspondence, we collectively concluded that the following very simple technique should lead to the desired objective:

The study of optical resonance, for example, that of the mercury atom (ref. 11, Chapter V), had shown that, in the presence of a magnetic field H_0, excitation with polarized light, or simply with a light beam directed in space, made it possible to obtain a selective excitation of the Zeeman sublevels of the excited state and that this selection still took place in a zero magnetic field[17]. Thus, in the case of the even isotopes of mercury, excitation by the 2537-Å line with polarization π leads solely to the sublevel $m=0$ of the excited state 6^3P_1, whereas excitation with circular polarization σ^+ or σ^- leads, respectively, to the sublevels $m=+1$ or $m=-1$ of this state. This selective excitation is reflected by the polarization of the resonance light emitted again when the excited atom is not perturbed during the short lifetime of the excited state ($\sim 10^{-7}$ sec). If, while maintaining a constant magnetic field H_0 which separates the Zeeman sublevels from the excited state, one applies perpendicular to this field a radio-frequency magnetic field, $H_1 \cos \omega t$, whose pulsation ω coincides with the Larmor frequency ω_0, magnetic resonance transitions are induced between the Zeeman sublevels of the excited state, and these transitions are manifested by a depolarization of the light emitted by optical resonance. (In the past, Fermi and Rasetti[18] had already applied an alternating magnetic field to excited atoms, but under conditions which did not correspond to a resonance phenomenon.) Therefore, the observation of the state of polarization of this light permits the optical detection of the magnetic resonance of excited states. We pointed out in the same note that, when the electron beam has a given direction, as in the experiment of Franck and Hertz[19], the excitation of atoms by electron impact also led to the emission of polarized spectral lines[20]; this proved that this mode of excitation also insured a selective excitation of the Zeeman sublevels of the excited states (alignment), and therefore that this should permit the optical detection of the radio-frequency resonances of these states through observation of the depolarization of the emission lines originating therefrom.

When Jean Brossel was applying the double-resonance method (it combines a magnetic resonance with an optical resonance) to the study of the 6^3P_1 state of the mercury atom, I showed, in an article in *Journal de Physique* of 1950[21], that the optical excitation of atoms with circularly polarized light made it possible to transfer the angular momentum carried bij the light to the atoms and thus to concentrate them in the ground state, either in the positive m sublevels or in the negative m sublevels (depending upon whether the light is σ^+ or σ^-) and that it was possible, by this optical pumping, to create an atomic orientation and also, due to the coupling between the electronic

magnetic moment and the nuclear spin, a nuclear orientation. In this manner, it should have been possible to obtain distributions very different from the Boltzmann distribution and thus to create conditions permitting the study of the return to equilibrium, either by relaxation or under the influence of a resonant field.

I must confess that, at that time, I had absolutely no knowledge of the slowness of the relaxation processes in the ground state, processes which take place in collisions with the wall or with the molecules of a foreign gas. Essentially, I had planned experiments on atomic beams in vacuum in order to avoid these relaxation effects. It was only later, in the course of the development of the experiments, that it became apparent that the relaxation processes on the walls are slow or can be substantially slowed down by suitable coatings[22], that the oriented ground states, insofar as they are orbital *S* states, are much less sensitive to collisions than the excited *P* states[23], and that, consequently, diamagnetic foreign gases can act as buffer gases. This observation later permitted considerable simplification of the experimental technique by working with vapors in a sealed container. When it was found, on the other hand, that the transverse relaxation times were of the same order of magnitude as the longitudinal times[24], this made it possible to obtain very narrow resonance lines and led to metrological applications which had not been suspected at the start[25].

In 1951, after finishing his pioneering work at M.I.T. on the excited state of the various mercury isotopes[26] and thus acquiring a knowledge of the techniques of Hertzian resonance, Jean Brossel returned to Paris. We then decided to organize a team of young research workers recruited from the students of the Écoles Normales Supérieures in order to develop systematically the optical methods of Hertzian resonance. The young people from this team are those who, in about a dozen theses, made personal and original contributions to the common work which is being honored today. In the meantime, the methods we had advocated and applied were picked up by a large number of foreign laboratories, leading to considerable improvements in the technique, which we, in turn, adopted and which were the source of great advances in the research work of our team.

The studies on the excited and ground states of the atoms occupied a primary place in the work of our team and led to a rich harvest of results: we collected numerous data on relaxation processes; from the position of the resonance lines, we were able to make precise measurements of Landé factors and intervals of fine and hyperfine structure and to deduce from them very

precise values of nuclear magnetic moments. We were led to the discovery of numerous phenomena related to high-order perturbations: multiple quantum transitions, effects of Hertzian coherence, demonstration of Hertzian resonance shifts under the influence of optical irradiation, and profound modification of the properties of an atom by the presence of a radio-frequency field. At the same time, with our techniques, foreign teams were achieving important results: measurement of nuclear quadrupole electric moments of alkali metal atoms, discovery of exchange collisions, displacement of hyperfine resonances by collisions with molecules of a foreign diamagnetic gas, and others.

In the course of the development of our research, we frequently had the satisfaction of seeing our predictions confirmed by experiments, but several times the experimental results were contrary to our predictions, thus creating problems whose solution led to advances that were as interesting as they were unexpected. The first piece of research carried out by our team was an example of this.

Indeed, resuming Brossel's experiments on the 6^3P_1 state of the mercury atom and adding an electric field to the magnetic field, Blamont studied the Stark effect of this state for the various even and odd isotopes[27] and discovered a narrowing of the magnetic-resonance curves as the density of mercury vapor increases (Fig. 1). The width of the magnetic-resonance curves of an excited state of an atom, extrapolated to zero amplitude of the radio-frequency field H_1 is indeed, as had been shown by Brossel, inversely proportional to the lifetime of this state, and the study of this width permits the measurement of this lifetime. This is a direct consequence of the uncertainty principle. The narrowing observed by Blamont thus appeared to contradict this principle, but Brossel found the explanation for this paradox: the phenomenon occurs because the transverse quantities (Hertzian coherences) are transmitted from atom to atom by processes of multiple scattering of optical-resonance photons. There is thus an «imprisonment of the Hertzian coherence» in the vapor, and this is manifested by the lengthening of the coherence time and by the narrowing of the resonance curves. Indeed, Mlle. Rollet's work, performed within our team[28], had just shown that the increasing depolarization of resonance light as a function of the mercury vapor density was due to this multiple scattering of photons and not to collisions, because the depolarization effect is more rapid in a pure isotope than in the natural isotope mixture (Fig. 2). Mlle. Guiochon proved the accuracy of Brossel's hypothesis[29] by showing that only the atoms of the same isotope in the mixture of the resonance cell

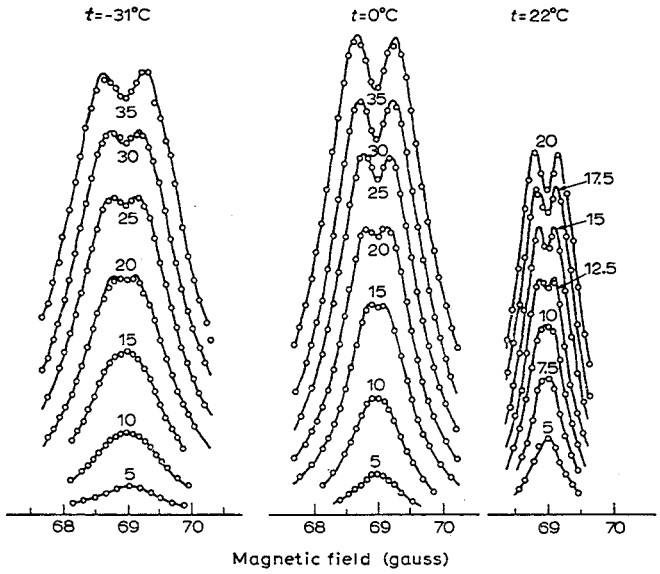


Fig. 1. Magnetic resonance curve networks of the 6^3P_1 level of the mercury atom. Each curve corresponds to a constant amplitude of the radio-frequency field H_1. The numbers indicate the relative values of these amplitudes in arbitrary units. Temperature t of the mercury drop determines the vapor density. (Guiochon, Blamont and Brossel[29])

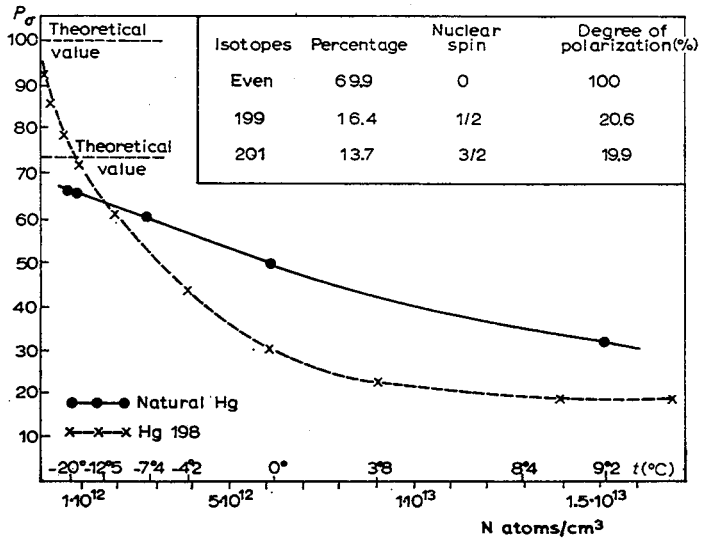


Fig. 2. Degree of polarization of optical resonance light at 2537 Å as a function of vapor density for natural mercury and for the pure ^{198}Hg isotope. (Rollet, Brossel and Kastler[28])

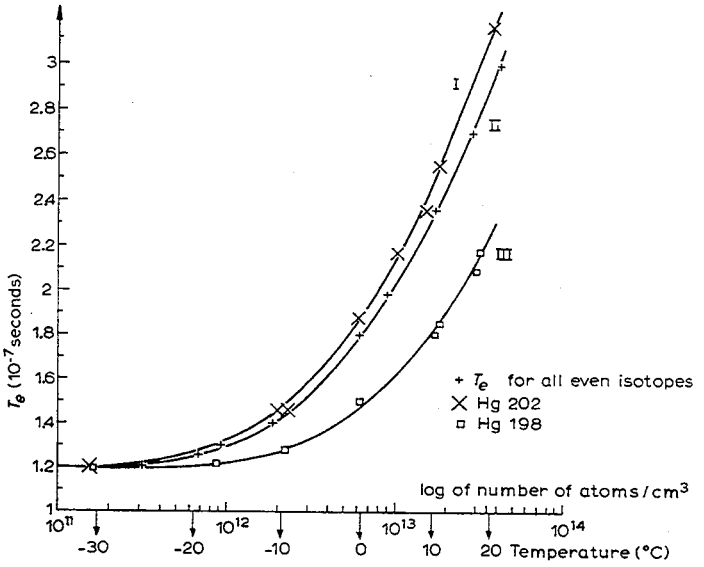


Fig. 3. Coherence time T_e deduced from the width of magnetic resonance curves as a function of the vapor density. Natural mercury in resonance cell. Different light sources: natural mercury, ^{202}Hg and ^{198}Hg. (Guiochon, Blamont and Brossel[29])

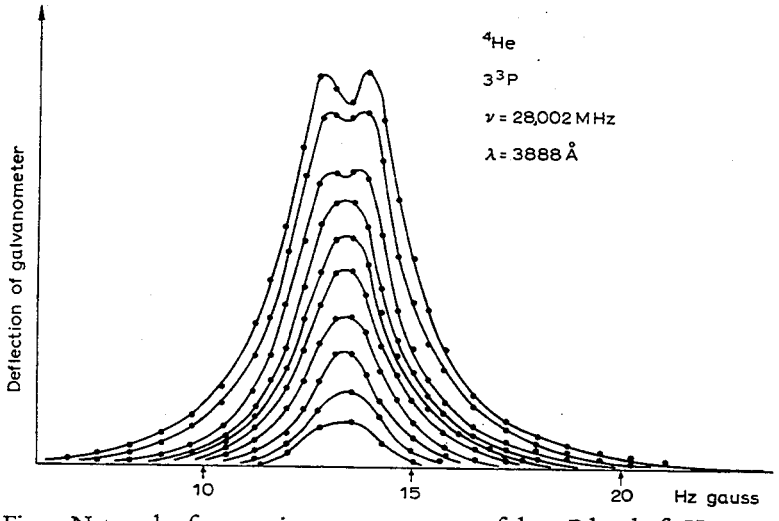


Fig. 4. Network of magnetic resonance curves of the $3\,^3P$ level of ^4He atom. (Descoubes[35])

produce the narrowing (Fig. 3). In his thesis, Jean-Pierre Barrat[30] developed the theory of coherent scattering and verified experimentally all the predictions of this theory. This was the first example of the study of an effect of Hertzian coherence between atomic states, and Barrat showed that these effects can be described in the formalism of the density matrix[31] which later proved to be very useful and fruitful in the study of other Hertzian coherence effects[32].

We charged Pebay-Peyroula with the task of testing the method of excitation by electron impact. In his thesis, he showed the fruitfulness of this technique[33] which later was developed by J. P. Descoubes, who combined it with the level crossing method[34]; this permitted him to analyze the fine and hyperfine structure of a large number of levels of ^{4}He and ^{3}He atoms[35] (Fig. 4).

The first tests of the optical pumping technique on a beam of sodium atoms[36], from the very first application of a radio-frequency field, led to the discovery of multiple quantum transitions[37]. Fig. 5 shows the first resonance

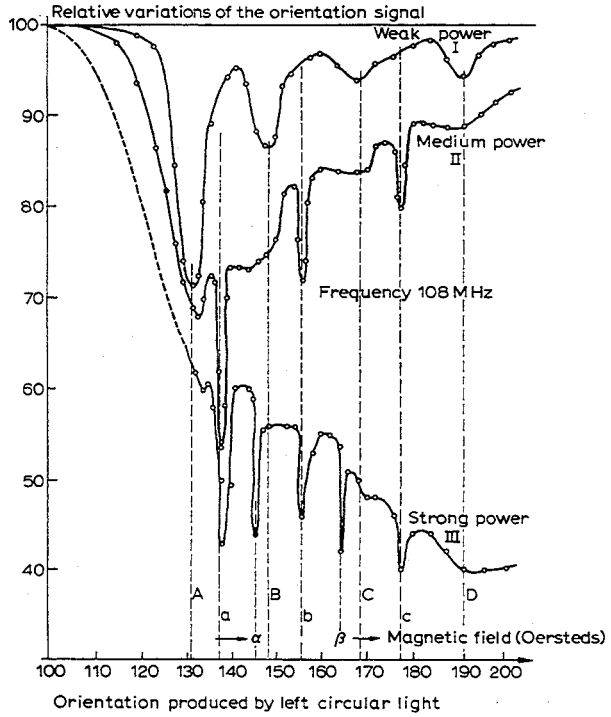


Fig. 5. Magnetic resonance curves of ^{23}Na at a constant frequency of 108 MHz and a variable field H_0 for increasing powers of the radio-frequency field. A, B, C, D, normal resonances $\Delta m = 1$; a, b, c, resonances with two quanta $\Delta m = 2$; α, β, resonances with three quanta $\Delta m = 3$. (Brossel, Cagnac and Kastler[37])

curves of the ^{23}Na atom with nuclear spin $I=3/2$, where, in a field of about 100 gauss, the ordinary Zeeman resonances are already widely separated by the decoupling effect between electronic and nuclear spin. This figure shows the narrow intermediate resonances corresponding to the transition $\Delta m=2$ induced by the absorption of two radio-frequency quanta.

The study of these multiple quantum transitions was systematically undertaken by J. M. Winter[38] who used them to erect a complete theory and was able to predict the existence of a new type of such transitions: in an atomic system possessing only two levels ($m=-\frac{1}{2}$ and $m=+\frac{1}{2}$), multiple quantum transitions are possible when the radiation field contains quanta of different polarization states and when the principle of conservation of energy and the principle of conservation of angular momentum can be satisfied at the same time. In addition, the theory predicts typical radiative broadenings and shifts as a function of the amplitude of the radio-frequency field, and the experiment has verified all predictions one by one.

Fig. 6 shows an example of multiple quantum transitions in the case of the ground state $I=\frac{1}{2}$ of the ^{199}Hg isotope.

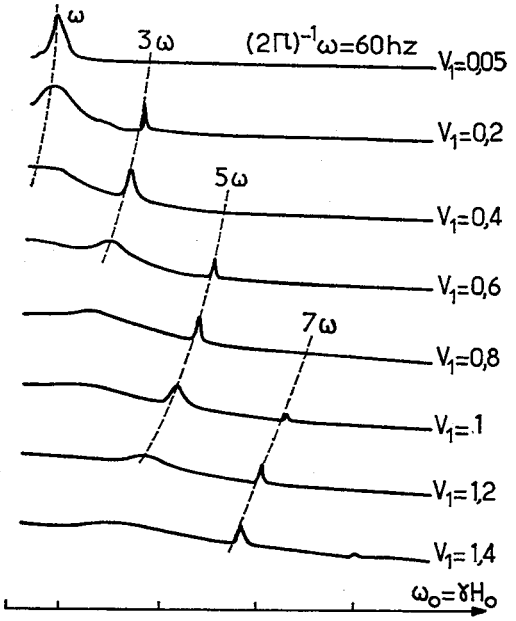


Fig. 6. Resonances with several quanta observed in the ground state. $I=\frac{1}{2}$ of ^{199}Hg; $\omega_0=n\omega$ with $n=1, 3, 5, 7$. From top to bottom, the curves correspond to increasing powers of radio frequency. V_1 indicates the voltage value measured across the terminals of the radio-frequency circuit. (Cohen-Tannoudji and Haroche)

As already noted, the discovery of paraffin coatings[22] and the effect of buffer gases[23] considerably facilitated the optical pumping of alkalis in the vapor phase and led to the discovery in the United States of exchange collisions[39,40]. In our team, Mme. M. A. Bouchiat-Guiochon used paraffin coatings to elucidate the relaxation mechanism on the container wall[41].

After various failures[42], Bernard Cagnac was the first successfully to obtain nuclear orientation in a low-density vapor (^{199}Hg and ^{201}Hg) by optical pumping and to use it in a study of the nuclear magnetic resonance of these atoms (Fig. 7); with others, he obtained high-precision measurements of nuclear magnetic moments[43]. Taking advantage of Franzen's elegant method of transients[44], he was also able to study relaxation processes due to collisions against the walls (Fig. 8), which led to very interesting problems of surface physics[45].

In similar fashion, J. C. Lehmann was able to orient the nuclei of the odd cadmium isotopes, to observe their magnetic resonance curves, and to measure precisely their nuclear magnetic moments[46], but he failed in attempts to orient the ^{67}Zn nuclei by pumping with the singlet resonance line despite the very high transition probability of this line. This failure caused him to analyze closely the process of nuclear orientation by optical pumping and to show that

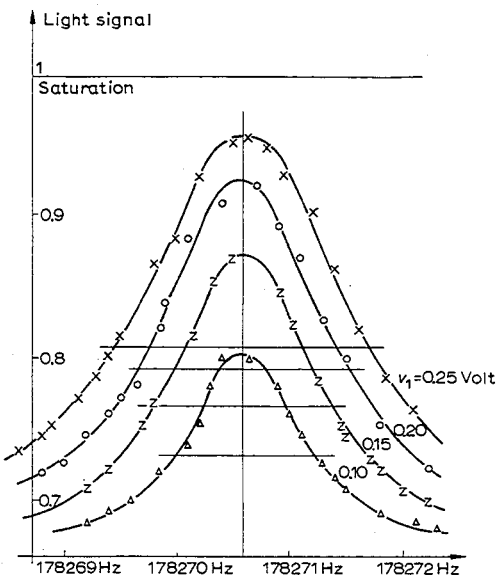


Fig. 7. Network of nuclear magnetic resonance curves of ^{199}Hg in the ground state. Fixed field H_0, variable frequency. (Cagnac[43])

this orientation does not take place during the light-absorption process, but is produced under the influence of the recoupling of the nuclear and electronic moments in the intermediate stage between absorption and reemission of light[47]. A detailed analysis of the process of nuclear orientation led Lehmann to the development of a method for measuring unresolved hyperfine intervals smaller than the natural width of the levels studied and to a successful application of this method to the odd cadmium isotopes.

We are thus led to a generalization of the Franck–Condon principle: «In a rapid process involving the electronic configuration (spectral transition, disorienting collision, or exchange collision), the position *and orientation* of the atomic nuclei remain unchanged.» The consequences of this principle were verified for disorienting collisions in the excited state by Omont and Faroux[48], and for exchange collisions in the ground state by Mme. Grossetête[49].

The cross-beam technique introduced by Hans Dehmelt[50] proved to be of considerable importance by making the behavior of the transverse macroscopic magnetic moment of a paramagnetic vapor accessible to optical detection (Fig. 9). Such a moment which precesses around the constant field H_0 produces a modulation of the absorption of a crossed beam perpendicular to the primary beam and to the field H_0. This modulation can be readily am-

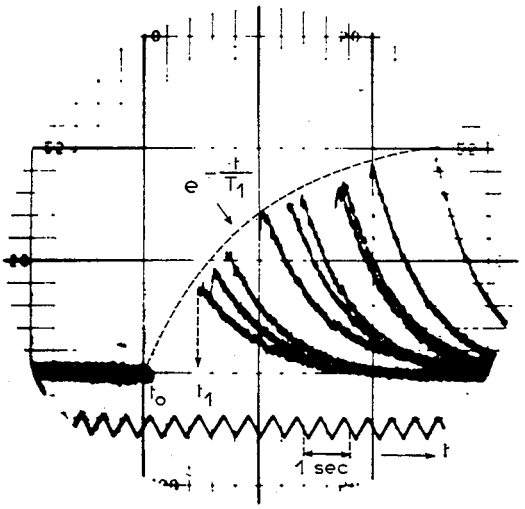


Fig. 8. Transient optical pumping curves of ^{199}Hg photographed on a cathodic oscilloscope screen. After a period of optical pumping, the atoms relax during a dark interval $t_1 - t_0$, and the pumping resumes. This time interval is modified from one curve to another. The dashed curve defines the relaxation exponential and permits measurement of the longitudinal relaxation time T_1. (Cagnac[43])

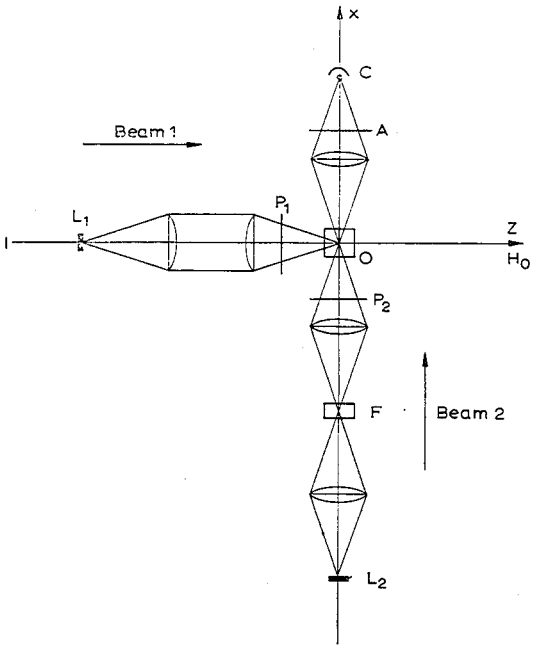


Fig. 9. Diagram of the apparatus used in Dehmelt's cross-beam method. Beam 1, pumping beam in circular light. Beam 2, detection beam whose intensity is modulated by the magnetic resonance; C, photodetector. (Cohen-Tannoudji[51])

plified, and its phase can be precisely determined by synchronous detection techniques.

Using ^{199}Hg as an example, C. Cohen-Tannoudji has shown the advantages inherent in this technique[51], either in studying steady-state resonance effects or in observing transient phenomena. Fig.10 shows how the technique of synchronous detection makes it possible to separate the components of the transverse moment, one of which is in phase and the other out of phase with the alternating field H_1 producing the resonance. The first of these components varies like a dispersion curve, and the second like an absorption curve. Fig.11 shows the transient signal obtained when the primary pumping beam and the radio-frequency field are interrupted at the same instant. The exponential decrease of the free precession of the transverse moment is then recorded as a function of time, and from this the transverse relaxation time is directly deduced. The technique of these transients can be combined with 90° and 180° pulse methods. Fig.12 shows the optical signals produced by these pulses.

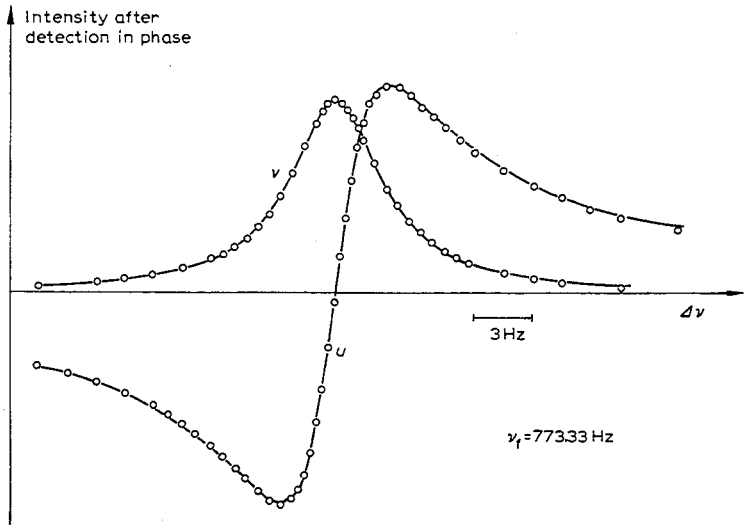


Fig.10. Nuclear magnetic resonance curves of ^{199}Hg observed by the modulation of a cross-beam. Synchronous detection permits the separation of components u and v of the transverse moment: u, component in phase with H_1; v, component out of phase with H_1. (Cohen-Tannoudji[51])

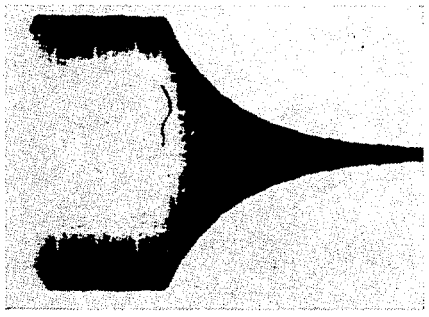
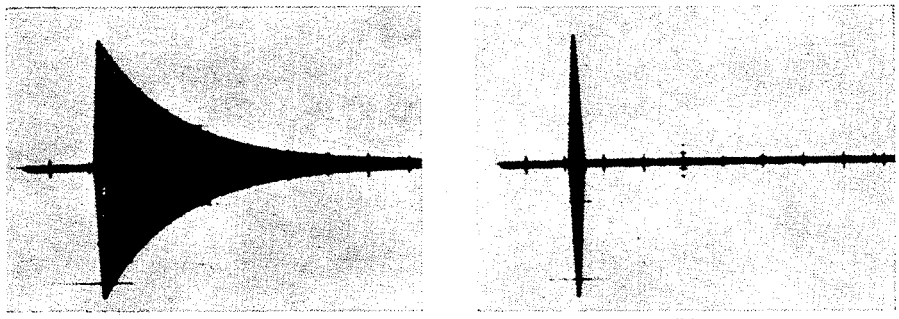


Fig.11. Transient signal of the modulation of the cross-beam: ^{199}Hg. Effect of sudden and simultaneous interruption of the pumping light beam and the radio-frequency field. Free decay of the transverse momentum. Exponential giving transverse relaxation time T_2 (Cohen-Tannoudji[51])

Cohen-Tannoudji used the refinement of these techniques to study the effect of energy shifts caused by a luminous irradiation[51] and predicted by the quantum theory of the optical pumping cycle[52]. He showed the existence of two kinds of such shifts: shifts due to real optical transitions which, due to the oscillation of the atom between the ground state and the excited state, produce



(a) (b)

Fig.12. Transient signals of the modulation of cross-beam: ^{199}Hg. (a) Effect of a 90° radio-frequency pulse; (b) effect of a 180° radio-frequency pulse. (Cohen-Tannoudji[51])

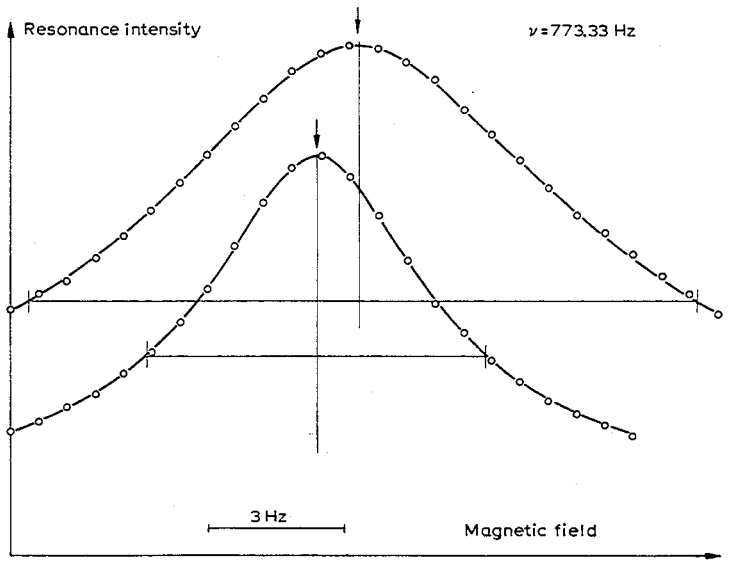


Fig.13. Nuclear magnetic resonance curves of ^{199}Hg observed by modulation of the cross-beam. For the upper curve, the intensity of the cross beam is five times greater than for the lower curve. The increase of light intensity widens the resonance curve and shifts its center. (Cohen-Tannoudji[51])

a mixture of Larmor precessions of the two states (Fig.13), and shifts due to virtual transitions produced by radiation which is not absorbed by the atom but which is close to an absorption frequency (Fig.14). In the latter case, the interaction between the atom and the radiation field is manifested by two complementary effects: the action of the atoms on the light produces a change

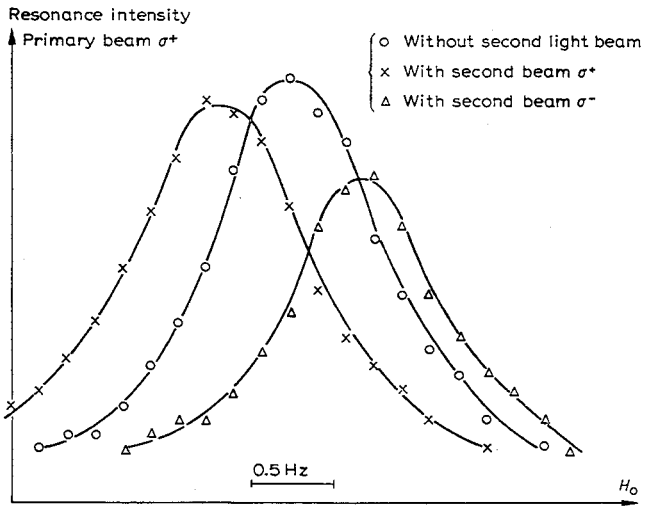


Fig. 14. Effect of virtual transitions. Shifts of the center of the ^{199}Hg nuclear magnetic resonance curve influenced by a second light beam (Cohen-Tannoudji[51])

in the velocity of propagation of the latter described by the abnormal dispersion curve, a phenomenon which has been known for a century; the action of the light on the atom produces a shift of the ground level of the atom. As a function of the deviation of the light frequency from the resonance frequency, the magnitude of the shift also varies like an abnormal dispersion curve (Fig. 15)[53].

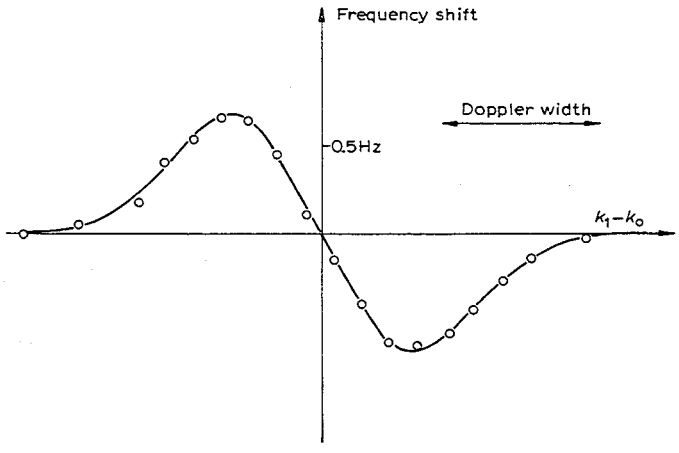


Fig. 15. Magnitude of shift as a function of $k_1 - k_0$. k_0, Center of optical absorption line of the atoms; k_1 center of acting radiation. k designs the wave number of the light. (Cohen-Tannoudji[51])

We should note that this latter type of shift was successfully amplified to a considerable extent by Russian physicists using the intense light of a ruby laser whose wavelength is close to that of a transition of the potassium atom[54].

Cohen-Tannoudji and his students applied themselves to a more extensive study of the interactions related to virtual absorptions and emissions of radio-frequency photons by an atom, interactions which give rise to new resonances whose characteristics are quite distinct from those of resonances corresponding to multiple quantum transitions described above[55].

Finally, when in the vicinity of zero value of H_0, an atom is placed in a non-resonant radio-frequency field, one observes a change of the Landé factor of the atom as a function of the intensity of this field[56], a change illustrated by Fig.16: all of these effects can be understood from a synthetic standpoint by studying the energy diagram of the total system «atom + radio-frequency photons»[57].

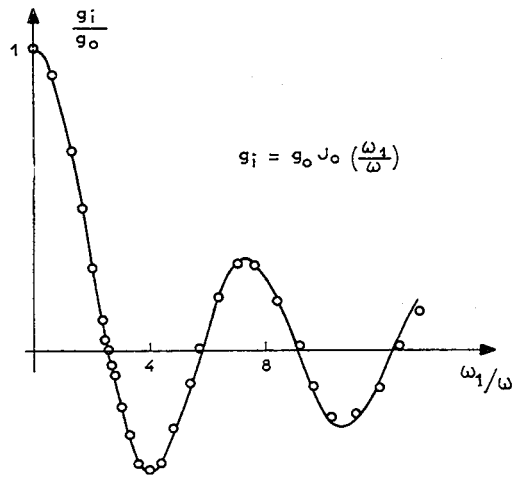


Fig. 16. Modification of the Landé factor of an atom (nuclear Zeeman effect of the ground state of ^{199}Hg) as a function of the intensity of radio-frequency field $H_1 \cos \omega t$ acting upon it. $\omega_1 = \lambda H_1$. (Cohen-Tannoudji and Haroche[56])

In conclusion, let us note that Jean Margerie[58] has shown that the optical methods of radio-frequency resonance can be transposed to paramagnetic ions and to *F* centers in solids and can yield valuable data on the structure of excited levels even in cases where the structure of these levels has not been spectrally resolved.

1. A. Sommerfeld, *Atombau und Spektrallinien*, 2nd ed., Vieweg, Braunschweig, 1922.
2. A. Rubinowicz, *Physik. Z.*, 19 (1918) 442, 465.
3. W. Hanle, *Naturwissenschaften*, 19 (1931) 375; R. Bär, *ibid.*, p. 463.
4. A. Kastler, *Compt. Rend.*, 193 (1931) 1075.
5. J. Cabannes, *J. Phys.*, 2 (1931) 381.
6. C. V. Raman and S. Bhagavantam, *Indian. J. Phys.*, 6 (1931) 353.
7. A. Kastler, *J. Phys.*, 2 (1931) 159.
8. R. Frisch, *Z. Physik*, 61 (1930) 626.
9. A. Kastler, *J. Phys.*, 4 (1933) 406.
10. C. Fuchtbauer, *Physik. Z.*, 21 (1920) 635.
11. A. C. G. Mitchell and M. W. Zemansky, *Resonance Radiation and Excited Atoms*, Cambridge University Press, London, 1934, p. 44.
12. A. Kastler, Thesis, *Ann. Phys. (Paris)*, 6 (1936) 663.
13. J. M. B. Kellog and S. Millman, *Rev. Mod. Phys.*, 18 (1946) 323.
14. W. E. Lamb Jr., Nobel Lecture 1955, Fine Structure of the Hydrogen Atom, in *Nobel Lectures Physics, 1942–1962*, Elsevier, Amsterdam, 1964, p. 286; *Rept. Progr. Phys.*, 14 (1951) 23.
15. F. Bitter, *Phys. Rev.*, 76 (1949) 833.
16. M. H. L. Pryce, *Phys. Rev.*, 77 (1950) 136.
17. W. Heisenberg, *Z. Physik*, 31 (1926) 617.
18. E. Fermi and F. Rasetti, *Z. Physik*, 33 (1925) 246.
19. J. Franck, Nobel Lecture 1925, Transformations of Kinetic Energy of Free Electrons into Excitation Energy of Atoms by Impacts, in *Nobel Lectures Physics 1922–1941*, Elsevier, Amsterdam, 1965, p. 98.
20. H. W. B. Skinner and E. T. S. Appleyard, *Proc. Roy. Soc. (London), Ser. A*, 117 (1928) 224.
21. A. Kastler, *J. Phys.*, 11 (1950) 255.
22. H. G. Robinson, E. S. Ensberg and H. G. Dehmelt, *Bull. Am. Phys. Soc.*, 3 (1958) 9.
23. J. Margerie, *Compt. Rend.*, 241 (1955) 865; P. Bender, *Thesis*, Princeton University, 1956; F. Hartmann, M. Rambosson, J. Brossel and A. Kastler, *Compt. Rend.*, 246 (1958) 1522.
24. C. Cohen-Tannoudji, J. Brossel and A. Kastler, *Compt. Rend.*, 245 (1957) 1027.
25. M. Arditi, Thesis, *Ann. Phys. (Paris)*, 5 (1960) 973; C. S. F. (Compagnie Générale de Télégraphie sans Fil, 97 Bd Haussmann, Paris, 8e), *Notice No.* 1602, 1964; L. Malnar and J. P. Mosnier, *Ann. Radioélectricité*, 16, No. 63 (1961); P. L. Bender, *Proc. 9th Ampère Colloquium, Pisa, 1960*, p. 621; A. L. Bloom, *Appl. Opt.*, 1 (1962) 61.
26. J. Brossel, Thesis, *Ann. Phys. (Paris)*, 7 (1952) 622; J. Brossel and F. Bitter, *Phys. Rev.*, 86 (1952) 308.
27. J. E. Blamont, Thesis, *Ann. Phys. (Paris)*, 2 (1957) 551.
28. N. Rollet, J. Brossel and A. Kastler, *Compt. Rend.*, 242 (1956) 240.
29. M. A. Guiochon, J. E. Blamont and J. Brossel, *Compt. Rend.*, 243 (1956) 1859; *J. Phys.*, 18 (1957) 99.
30. J.-P. Barrat, Thesis, *J. Phys.*, 20 (1959) 541, 633, 657.
31. J.-P. Barrat, *Proc. Roy. Soc. (London), Ser. A*, 263 (1961) 371.

32. C. Cohen-Tannoudji, in *Quantum Electronics*, Vol. 2, Columbia University Press, New York, 1961, p. 114.
33. J. C. Pebay-Peyroula, Thesis, *J. Phys.*, 20 (1959) 669, 721.
34. P. A. Franken, *Phys. Rev.*, 121 (1961) 508.
35. J. P. Descoubes, *Compt. Rend.*, 259 (1964) 327, 3733, 261 (1965) 916; J. P. Descoubes, *Thesis*.
36. J. Brossel, A. Kastler and J. M. Winter, *J. Phys.*, 13 (1952) 668.
37. J. Brossel, B. Cagnac and A. Kastler, *Compt. Rend.*, 237 (1953) 984; *J. Phys.*, 15 (1954) 6.
38. J. M. Winter, Thesis, *Ann. Phys. (Paris)*, 4 (1959) 745.
39. H. G. Dehmelt, *Phys. Rev.*, 109 (1958) 381.
40. P. A. Franken, R. Sands and J. Hobart, *Phys. Rev. Letters*, 1 (1958) 316; R. Novick and H. E. Peters, *ibid.*, pp. 54, 152.
41. M. A. Bouchiat, *Thesis*, Publications Scientifiques du Ministère de l'Air, 1965, No. N. T. 146; M. A. Bouchiat and J. Brossel, *Phys. Rev.*, 147 (1966) 41.
42. F. Bitter, R. F. Lacey and B. Richter, *Rev. Mod. Phys.*, 25 (1953) 174.
43. B. Cagnac, Thesis, *Ann. Phys. (Paris)*, 6 (1960) 467; J. C. Lehmann and R. Barbé, *Compt. Rend.*, 257 (1963) 3152.
44. W. Franzen, *Phys. Rev.*, 115 (1959) 850.
45. C. Cohen-Tannoudji, *J. Phys.*, 24 (1963) 653; C. Cohen-Tannoudji and J. Brossel, *Compt. Rend.*, 258 (1964) 6119; J. C. Lehmann and J. Brossel, *ibid.*, 262 (1966) 624.
46. J. C. Lehmann, *Thesis*, Paris, 1966; J. C. Lehmann and J. Brossel, *Compt. Rend.*, 258 (1964) 869; M. Leduc and J. C. Lehmann, *ibid.*, 262 (1966) 736.
47. J. C. Lehmann, *J. Phys.*, 25 (1964) 809.
48. A. Omont, *J. Phys.*, 26 (1965) 26; J. P. Faroux and J. Brossel, *Compt. Rend.*, 261 (1965) 3092, 262 (1966) 41; J. P. Faroux, *ibid.*, 262 (1966) 1385.
49. F. Grossetête, *J. Phys.*, 25 (1964) 383; *Compt. Rend.*, 259 (1964) 3211, 260 (1965) 3327.
50. H. G. Dehmelt, *Phys. Rev.*, 105 (1957) 1924; W. E. Bell and A. L. Bloom, *ibid.*, 107 (1957) 1559.
51. C. Cohen-Tannoudji, Thesis, *Ann. Phys. (Paris)*, 7 (1962) 423, 469.
52. J. P. Barrat and C. Cohen-Tannoudji, *Compt. Rend.*, 252 (1961) 93; *J. Phys.*, 22 (1961) 329.
53. See ref. 51, and A. Kastler, *J. Opt. Soc. Am.*, 53 (1963) 902.
54. E. B. Alexandrov, A. M. Bonch-Bruevich, N. N. Konstin and V. A. Khodovoi, *J. Exptl. Theoret. Phys. (U. S. S. R.) Letters*, 3 (1966) 53.
55. N. Polonsky and C. Cohen-Tannoudji, *J. Phys.*, 26 (1965) 409; *idem*, *Compt. Rend.*, 260 (1965) 5231, 261 (1965) 369; C. Cohen-Tannoudji and S. Haroche, *ibid.*, 261 (1965) 5400.
56. C. Cohen-Tannoudji and S. Haroche, *Compt. Rend.*, 262 (1966) 268.
57. C. Cohen-Tannoudji and S. Haroche, *Compt. Rend.*, 262 (1966) 37.
58. J. Margerie, *Thesis*, Publications Scientifiques du Ministère de l'Air, 1966, No. N. T. 155; J. Brossel and J. Margerie, in W. Low (Ed.), *Paramagnetic Resonance*, Vol. 2, Academic Press, New York, 1963, p. 535; N. V. Karlov, J. Margerie and Merle d'Aubigné, *J. Phys.*, 24 (1963) 717; see also A. Kastler, in W. S. C. Chang (Ed.). *Lasers and Applications*, Ohio State University Press, Columbus, 1963.

Bibliography

H. Kopfermann, *Kernmomente*, 2nd ed., chapter 24, Die Doppelresonanz Methode, Akad. Verlagsgesellschaft, Frankfurt, 1956.

A. Kastler, *J. Opt. Soc. Am.*, 47 (1957) 460; *Physikertagung*, Mosbach, Wiesbaden, 1960, p. 62; *Proc. Ampère Colloquium, Eindhoven*, North-Holland, Amsterdam, 1962; *J. Opt. Soc. Am.*, 53 (1963) 902.

G. W. Series, *Rept. Progr. Phys.*, 22 (1959) 280.

T. Skalínski, *Rendiconti Scuola Internazionale de Fisica, Varenne, 1960*, Academic Press, London, 1962, p. 212.

G. zu Putlitz, *Ergeb. Exakt. Naturw.*, 37 (1965) 105.

J. Brossel, *Yearbook Phys. Soc. (London)*, (1960) 1.

J. Brossel, in *Quantum Electronics*, Vol. 1, Columbia University Press, New York, 1960, p. 81; in *Quantum Electronics*, Vol. 2, Columbia University Press, New York, 1961, p. 95; *Proc. Les Houches Conference on Optics and Quantum Electronics*, Gordon and Breach, New York, 1964, p. 187.

G. V. Skrotsky and I. G. Izioumova, *Soviet Phys. Usp.*, 4 (1961) 177.

R. A. Bernheim, *Optical Pumping: An Introduction*, Benjamin, New York, 1965.

A. Kastler, in *Proc. Zeeman Centennial Conference*, Netherlands Phys. Soc., Amsterdam, 1965, p. 73.

C. Cohen-Tannoudji and A. Kastler, in *Progress in Optics*, Vol. 5, North-Holland, Amsterdam, 1966, p. 1.

A. Kastler, in *Proc. Ampère Colloquium, Ljubljana, 1966*, North-Holland, Amsterdam, 1967.

A. L. Bloom, *Sci. Am.*, 203 (1960) 72.

R. de Zafra, *Am. J. Phys.*, 28 (1960) 646.

T. R. Carver, *Science*, 141 (1963) 599.

R. Benumof, *Am. J. Phys.*, 33 (1965) 151.

Biography

Alfred Kastler was born in Guebwiller in Alsace on May 3, 1902. He followed his early studies at the school in his native town, and continued at the Oberrealschule of Colmar, which became the Lycée Bartholdi in 1918, when Alsace was returned to France.

He entered the École Normale Supérieure in 1921, and left in 1926 to teach in a lycée. He taught for 5 years, first in the Mulhouse lycée, then in those of Colmar and Bordeaux. The next stage of his career was in higher education: assistant at the Bordeaux Faculty of Science from 1931 to 1936, lecturer at Clermont-Ferrand from 1936 to 1938, professor at Bordeaux from 1938 to 1941. In 1941, in the midst of the German occupation, Georges Bruhat asked him to come to Paris to help him in establishing physics teaching at the École Normale Supérieure. The post was provisional, but was confirmed by the allocation of a chair in a personal capacity at the Paris Faculty of Sciences in 1952.

His mathematics teachers at the Colmar Lycée, Fröhlich from Bavaria and Edouard Greiner from Alsace, were the first to awaken his interest in science. This predilection became consolidated in the special mathematics class held by Mahuet and Brunold, who helped Kastler to gain entry to the École Normale Supérieure by the side entrance, so to speak. In the stimulating and friendly atmosphere of this college, the teacher Eugène Bloch (who came from the upper Rhine and who subsequently disappeared without trace in Auschwitz) initiated his students into the concepts of Bohr's atom and quantum physics, and drew Kastler's attention to Sommerfeld's book on atomic structure and spectral lines. This book introduced him to the principle of the conservation of momentum applied by A. Rubinowicz to the exchange of energy between atoms and radiation. This principle was to guide the whole of Kastler's research, beginning with his thesis up to the most recent investigations of the Parisian team.

Alfred Kastler was in 1931 appointed assistant to Pierre Daure, professor at the Bordeaux Faculty of Science. His teaching duties were then less onerous, and Kastler was able to devote all his free time to research, aided by Professor

Daure who initiated him into experimental spectroscopy. For many years, he worked in the field of optical spectroscopy, particularly on atomic fluorescence and Raman spectroscopy. [In 1937 he became interested in the luminescence of sodium atoms in the upper atmosphere; after establishing that the D line of the twilight sky could be absorbed by sodium vapour, and after some studies at Abisko where twilight is prolonged, he was able to demonstrate in cooperation with his colleague Jean Bricard, that this line is polarized, as it must be if the emission mechanism is one of optical resonance produced by solar radiation.]

During the years of the occupation, French scientists were virtually isolated from the outside world. In 1945, it was possible to send pupils to other western countries, so that they could bring their knowledge of the most recent developments in scientific progress up to date. Among them was Jean Brossel, who returned in 1951 in possession of a mass of information gained under Francis Bitter at M.I.T.

Under the influence of Gorter, Rabi had very successfully applied certain methods to the investigation of atoms in their fundamental state. In 1949, Bitter suggested extending these same methods to the excited states of atoms. Brossel and Kastler together then proposed the «double resonance method», which combines optical resonance with magnetic resonance.

While Brossel was at M.I.T., between 1949 and 1951, he carried out pioneer work along these lines on the excited state of the mercury atom. At the same time, Kastler was supplementing the method by the technique of «optical pumping», which makes it possible to apply «optical methods for studying the microwave resonances» to the fundamental states of atoms.

After 1951, Kastler worked in collaboration with Jean Brossel in Paris to perfect all these methods. Among the young men and women at the École Normale, which nurtures the intellectual elite, they found their research workers. Their theses represent the various stages in their collective work which has been awarded the Nobel prize, and of which some account is given in Kastler's Nobel lecture.

Kastler taught as Francqui Professor at the University of Louvain during the year 1953–1954, he hold honorary doctorates from the University of Louvain (1955), Pisa (1960), and Oxford (1966), and he was decorated by the University of Liège. The French and Polish Societies of Physics and the American Society of Optics have elected Kastler to honorary memberships. In 1962, the latter society awarded him the first Mees medal bearing the inscription «Optics transcends all boundaries». In 1954, the British Physical Society

awarded him the prize commemorating Fernand Holweck, who disappeared tragically in 1941. Kastler was made a member of the Royal Flemish Academy of Belgium in 1954, and of the Paris Academy of Sciences in 1964; in 1965, the National Centre for Scientific Research awarded him their gold medal, at the same time as his friend and colleague Louis Néel.

In December 1924 Kastler married Elise Cosset, a former pupil of the École Normale Supérieure. By working as a history teacher in secondary schools she made it possible for her husband to devote to research all the leisure time left to him by his own teaching duties. They have three children: Daniel, born in 1926, Mireille born in 1928, and Claude-Yves born in 1936. They have all married, there are now six grandchildren, whose ages range from 14 years to 10 months. Daniel is a Professor of Physics at the Faculty of Science in Marseilles, he is working on theoretical physics problems; Mireille is an ophthalmologist in Paris, and Claude-Yves teaches Russian at the Arts Faculty in Grenoble.

Physics 1967

209

HANS ALBRECHT BETHE

«for his contributions to the theory of nuclear reactions, especially his discoveries concerning the energy production in stars»

Physics 1967

Presentation Speech by Professor O. Klein, member of the Swedish Academy of Sciences

Your Majesty, Your Royal Highnesses, Ladies and Gentlemen.

This year's Nobel Prize in Physics – to professor Hans A. Bethe – concerns an old riddle. How has it been possible for the sun to emit light and heat without exhausting its source not only during the thousands of centuries the human race has existed but also during the enormously long time when living beings needing the sun for their nourishment have developed and flourished on our earth thanks to this source? The solution of this problem seemed even more hopeless when better knowledge of the age of the earth was gained. None of the energy sources known of old could come under consideration. Some quite unknown process must be at work in the interior of the sun. Only when radioactivity, its energy generation exceeding by far any known fuel, was discovered, it began to look as if the riddle might be solved. And, although the first guess that the sun might contain a sufficient amount of radioactive substances soon proved to be wrong, the closer study of radioactivity would by and by open up a new field of physical research in which the solution was to be found.

While ordinary physics and chemistry could be led back to the behaviour of the electrons which form the outer part of atoms, the new field is concerned with their innermost part, the atomic nucleus. Its discoverer, Rutherford, called it the newer alchemy because nuclear reactions, in contrast to chemical reactions, usually lead to transmutations of the chemical elements – what alchemists wished to produce but could not by their means – the reaction energy being there some million times greater than in chemical reactions.

It soon became clear that the proton, the nucleus of the hydrogen atom, is a common building stone of all atomic nuclei. It is electrically charged. The other building stone, the neutron, being electrically neutral as indicated by its name, was discovered in 1932, twenty-one years later than the nucleus itself. And, in spite of important progress during those years, it may be said that from then on nuclear physics had really started. At that time it was already apparent that Bethe belonged to the small group of young theoretical physicists who through skill and knowledge were particularly qualified for tackling the

many theoretical problems turning up in close connection with the rapidly appearing experimental discoveries. The centre of these problems was to find the properties of the force that keeps the protons and neutrons together in the nucleus, the counterpart of the electric force which binds the atomic electrons to the nucleus. Bethe's contributions to the solution of these problems have been numerous and are still continuing. They put him clearly in the first row among the workers in this field–as in several other fields. Moreover, about the middle of the thirties he wrote, partly alone, partly together with some colleagues, what nuclear physicists at the time used to call the Bethe bible, a penetrating review of about all that was known of atomic nuclei, experimental as well as theoretical.

This extensive and profound knowledge of his regarding atomic nuclei together with a rare gift of rapidly grasping the essence of a physical problem and finding ways of solving it explains that Bethe could so swiftly do the work awarded by the Nobel Prize. He started his work after a conference taking place in Washington in March 1938 and the paper containing a thorough description of it was delivered for print at the beginning of September the same year. During that conference and afterwards he seems also to have acquired the necessary astrophysical knowledge. This knowledge depended mainly on a pioneer work by Eddington from the year 1926, according to which the innermost part of the sun is a hot gas mainly consisting of hydrogen and helium. Owing to the high temperature, about 20 million degrees,–these atoms being dissolved into electrons and nuclei–, the mixture, despite the high density – about 80 times that of water – really behaves like a gas. The amount of energy generation necessary to maintain this state was known from measurements of the radiation falling on the earth. Taken as a whole it is enormous, but very slow as compared to the size of the sun. An ordinary 60-Watt electric bulb would correspond to about 300 tons average sun matter. This very slow burning together with the very high energy release from a given weight of fuel gives this source the high durability required by geology and the long existence of life on the earth.

Before coming to the nuclear processes, which according to Bethe's work are definitely the source of the energy generation of the sun and similar stars, a few words should be said about two questions which naturally present themselves in this connection. Why are these nuclear processes so slow in the sun when they are so fast in atomic reactors, not to mention atomic bombs? And why are they non-existent under ordinary conditions? The answer is that nuclei are protected against other nuclei by the repulsion due to their electric

charges together with the extremely small range of the nuclear force–which is about as small relative to a midget as a midget is to the sun–implying that a proton must have an extremely high velocity in order to come so close to another nucleus as is necessary for a nuclear reaction to take place. If it were not for the quantum-mechanical tunnel effect studied very closely in this connection by Gamow–who must be considered the main forerunner of Bethe with respect to the application of nuclear physics to astronomy–even the velocities of the protons at the high temperature of the sun would not be able to produce any such processes. But through this effect the required slow reactions do occur. The case of the atomic reactors is different, because the reactions are there produced by neutrons, which having no charge are not stopped by the electric charge of the nuclei. Fortunately neutrons are short-lived and therefore extremely rare under ordinary circumstances and also in the sun.

Even when Bethe started his work on the energy generation in stars there were important gaps in the knowledge about nuclei which made the solution of the problem very difficult. And it was by a remarkable combination of underdeveloped theory and incomplete experimental evidence, under repeated comparison of his conclusions with their astronomical consequences, that he succeeded in establishing the mechanism of energy generation in the sun and similar stars so well that only minor corrections were needed when many years later the required experimental knowledge had made considerable progress and when, moreover, electronic computers had become available for the numerical calculations.

A very important part of his work resulted in eliminating a great number of thinkable nuclear processes under the conditions at the centre of the sun, after which only two possible processes remained. The simplest of them begins with two protons colliding and forming a nucleus of heavy hydrogen, the surplus of electric charge vanishing in the form of a positive electron. After capturing a few more protons the result of the process is the formation of a helium nucleus from four protons. Thereby the energy release from a given weight of hydrogen is nearly 20 million times greater than that produced by burning the same weight of carbon into carbon dioxide. The second process is more complicated. It requires the presence of carbon which, however, will practically not be consumed but acts as a catalyst, the result being the same as in the former process. It should be mentioned that the first process had been proposed a few years earlier by Atkinson and later discussed by von Weizsäcker, who also considered the second process independently of and at about the same time as Bethe. But none of them had attempted a thorough

analysis of these and other thinkable processes necessary to make it reasonably certain that these processes, and only these, are responsible for the energy generation in the sun and similar stars.

Bethe's work constitutes since many years a main foundation for the great development which has taken place of the knowledge of the interior of the sun and the stars. During recent years it has obtained a new actuality through a promising attempt made by a group of astrophysicists to understand what happens when a star has used up its hydrogen, thereby throwing new light on another old riddle, that of the origin of the chemical elements.

Professor Bethe. You may have been astonished that among your many contributions to physics, several of which have been proposed for the Nobel Prize, we have chosen one which contains less fundamental physics than many of the others and which has taken only a short part of your long time in science. This, however, is quite in agreement with the rules of the Nobel Prize and does not imply that we are not highly impressed by the role you have played in so many parts of the development of physics ever since you started doing research some forty years ago. On the other hand your solution of the energy source of stars is one of the most important applications of fundamental physics in our days, having led to a deepgoing evolution of our knowledge of the universe around us. On behalf of the Royal Swedish Academy of Sciences I extend to you the most hearty congratulations. And now I have the privilege to ask you to receive the Nobel Prize for Physics from the hands of His Majesty the King.

215

H. A. BETHE

Energy production in stars

Nobel Lecture, December 11, 1967

History

From time immemorial people must have been curious to know what keeps the sun shining. The first scientific attempt at an explanation was by Helmholtz about one hundred years ago, and was based on the force most familiar to physicists at the time, gravitation. When a gram of matter falls to the sun's surface it gets a potential energy

$$E_{\text{pot}} = -GM/R = -1.91 \cdot 10^{15} \text{ erg/g} \qquad (1)$$

where $M = 1.99 \cdot 10^{33}$ g is the sun's mass, $R = 6.96 \cdot 10^{10}$ cm its radius, and $G = 6.67 \cdot 10^{-8}$ the gravitational constant. A similar energy was set free when the sun was assembled from interstellar gas or dust in the dim past; actually somewhat more, because most of the sun's material is located closer to its center, and therefore has a numerically larger potential energy. One-half of the energy set free is transformed into kinetic energy according to the well-known virial theorem of mechanics. This will permit us later to estimate the temperature in the sun. The other half of the potential energy is radiated away. We known that at present the sun radiates

$$\varepsilon = 1.96 \text{ erg/g sec} \qquad (2)$$

Therefore, if gravitation supplies the energy, there is enough energy available to supply the radiation for about 10^{15} sec which is about 30 million years.

This was long enough for nineteenth century physicists, and certainly a great deal longer than man's recorded history. It was not long enough for the biologists of the time. Darwin's theory of evolution had just become popular, and biologists argued with Helmholtz that evolution would require a longer time than 30 million years, and that therefore his energy source for the sun was insufficient. They were right.

At the end of the 19th century, radioactivity was discovered by Becquerel and the two Curie's who received one of the first Nobel prizes for this discovery. Radioactivity permitted a determination of the age of the earth, and more

recently, of meteorites which indicate the time at which matter in the solar system solidified. On the basis of such measurements the age of the sun is estimated to be 5 milliards of years, within about 10%. So gravitation is not sufficient to supply its energy over the ages.

Eddington, in the 1920's, investigated very thoroughly the interior constitution of the sun and other stars, and was much concerned about the sources of stellar energy. His favorite hypothesis was the complete annihilation of matter, changing nuclei and electrons into radiation. The energy which was to be set free by such a process, if it could occur, is given by the Einstein relation between mass and energy and is

$$c^2 = 9 \cdot 10^{20} \text{ erg/g} \tag{3}$$

This would be enough to supply the sun's radiation for 1500 milliards of years. However nobody has ever observed the complete annihilation of matter. From experiments on earth we know that protons and electrons do not annihilate each other in 10^{30} years. It is hard to believe that the situation would be different at a temperature of some 10 million degrees such as prevails in the stars, and Eddington appreciated this difficulty quite well.

From the early 1930's it was generally assumed that the stellar energy is produced by nuclear reactions. Already in 1929, Atkinson and Houtermans[1] concluded that at the high temperatures in the interior of a star, the nuclei in the star could penetrate into other nuclei and cause nuclear reactions, releasing energy. In 1933, particle accelerators began to operate in which such nuclear reactions were actually observed. They were found to obey very closely the theory of Gamow, Condon and Gurney, on the penetration of charged particles through potential barriers. In early 1938, Gamow and Teller[2] revised the theory of Atkinson and Houtermans on the rate of «thermonuclear» reactions, *i.e.* nuclear reactions occurring at high temperature. At the same time, von Weizsäcker[3] speculated on the reactions which actually might take place in the stars.

In April 1938, Gamow assembled a small conference of physicists and astrophysicists in Washington, D.C. This conference was sponsored by the Department of Terrestrial Magnetism of the Carnegie Institution. At this Conference, the astrophysicists told us physicists what they knew about the internal constitution of the stars. This was quite a lot, and all their results had been derived without knowledge of the specific source of energy. The only assumption they made was that most of the energy was produced «near» the center of the star.

Properties of Stars

The most easily observable properties of a star are its total luminosity and its surface temperature. In relatively few cases of nearby stars, the mass of the star can also be determined.

Fig.1 shows the customary Hertzsprung–Russell diagram. The luminosity, expressed in terms of that of the sun, is plotted against the surface temperature, both on a logarithmic scale. Conspicuous is the main sequence, going from upper left to lower right, *i.e.* from hot and luminous stars to cool and faint ones. Most stars lie on this sequence. In the upper right are the Red Giants, cool but brilliant stars. In the lower left are the White Dwarfs, hot but faint. We shall be mainly concerned with the main sequence. After being assembled, by gravitation, stars spend the most part of their life on the main sequence, then develop into red giants, and in the end, probably into white dwarfs.

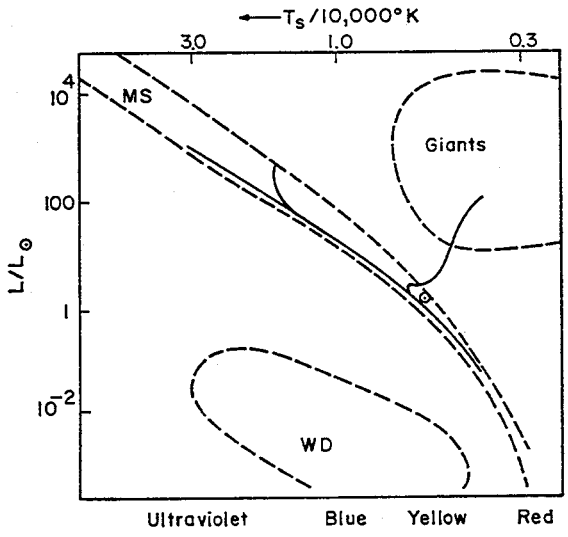


Fig.1. Hertzsprung–Russell diagram. From E.E.Salpeter, in *Apollo and the Universe*, Science Foundation for Physics, University of Sydney, Australia, 1967.

The figure shows that typical surface temperatures are of the order of 10^4 °K. Fig. 2 gives the relation between mass and luminosity in the main sequence. At the upper end, beyond about 15 sun masses, the mass determinations are uncertain. It is clear, however, that luminosity increases rapidly with mass.

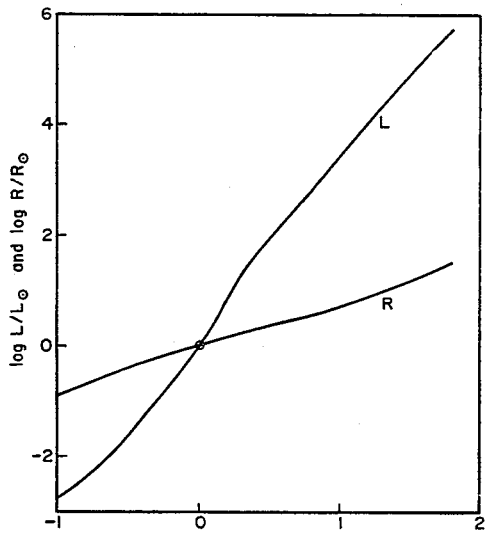


Fig. 2. Luminosity and radius of stars *vs.* mass. Abscissa is $\log M/M_\odot$. Data from C. W. Allen, *Astrophysical Quantities*, Athlone Press, 1963, p. 203. The curve for $\log L/L_\odot$ holds for all stars, that for $R/R_\odot$ only for the stars in the main sequence. The symbol $\odot$ refers to the sun.

For a factor of 10 in mass, the luminosity increases by a factor of about 3000, hence the energy production per gram is about 300 times larger.

To obtain information on the interior constitution of the stars, astrophysicists integrate two fundamental equations. Pioneers in this work have been Eddington, Chandrasekhar and Strömgren. The first equation is that of hydrostatic equilibrium

$$\frac{dP}{dr} = -GM(r)\frac{\varrho(r)}{r^2} \tag{4}$$

in which P is the pressure at distance r from the center, ϱ is the density and $M(r)$ is the total mass inside r. The second equation is that of radiation transport

$$\frac{1}{\varkappa\varrho}\frac{d}{dr}\left(\tfrac{1}{3}acT^4\right) = -\frac{L(r)}{4\pi r^2} \tag{5}$$

Here $\varkappa$ is the opacity of the stellar material for black-body radiation of the local temperature T, a is the Stefan–Boltzmann constant, and $L(r)$ is the flux of radiation at r. The value of L at the surface R of the star is the luminosity.

In the stars we shall discuss, the gas obeys the equation of state

$$P = RT\varrho/\mu \tag{6}$$

where R is the gas constant, while μ is the mean molecular weight of the stellar material. If X, Y and Z are respectively concentrations by mass of hydrogen, helium and all heavier elements, and if all gases are fully ionized, then

$$\mu^{-1} = 2X + \tfrac{3}{4}Y + \tfrac{1}{2}Z \tag{7}$$

In all stars except the very oldest ones, it is believed that Z is between 0.02 and 0.04; in the sun at present, X is about 0.65, hence $Y = 0.33$ and $\mu = 0.65$. In many stars the chemical composition, especially X and Y, vary with position r. The opacity is a complicated function of Z and T, but in many cases it behaves like

$$\varkappa = C\varrho\, T^{-3.5} \tag{8}$$

where C is a constant.

The integration of (4) and (5) in general requires computers. However an estimate of the central temperature may be made from the virial theorem which we mentioned in the beginning. According to this, the average thermal energy per unit mass of the star is one-half of the average potential energy. This leads to the estimate of the thermal energy per particle at the center of the star,

$$k\, T_c = \alpha\, \mu\, G\, H\, M/R \tag{9}$$

in which H is the mass of the hydrogen atom, and α is a constant whose magnitude depends on the specific model of the star but is usually about 1 for main sequence stars. Using this value, and (1), we find for the central temperature of the sun

$$T_{6c} = 14 \tag{10}$$

where T_6 denotes the temperature in millions of degrees, here and in the following. A more careful integration of the equations of equilibrium by Demarque and Percy[4] gives

$$T_{c6} = 15.7;\ \varrho_c = 158\ \text{g/cm}^3 \tag{11}$$

Originally Eddington had assumed that the stars contain mainly heavy elements, from carbon on up. In this case $\mu = 2$ and the central temperature is increased by a factor of 3, to about 40 million degrees; this led to contradictions with the equation of radiation flow, (5), if the theoretical value of the

opacity was used. Strömgren pointed out that these contradictions can be resolved by assuming the star to consist mainly of hydrogen, which is also in agreement with stellar spectra. In modern calculations the three quantities X, Y, Z, indicating the chemical composition of the star, are taken to be parameters to be fixed so as to fit all equations of stellar equilibrium.

Thermonuclear Reactions

All nuclei in a normal star are positively charged. In order for them to react they must penetrate each others Coulomb potential barrier. The wave mechanical theory of this shows that in the absence of resonances, the cross section has the form

$$\sigma(E) = \frac{S(E)}{E} \exp\left(-\sqrt{\frac{E_G}{E}}\right) \tag{12}$$

where E is the energy of the relative motion of the two colliding particles, $S(E)$ is a coefficient characteristic of the nuclear reaction involved and

$$E_G = 2M(\pi Z_0 Z_1 e^2/\hbar)^2 = (2\pi Z_0 Z_1)^2 E_{\text{Bohr}} \tag{13}$$

Here M is the reduced mass of the two particles, Z_0 and Z_1 their charges, and E_{Bohr} is the Bohr energy for mass M and charge 1. (13) can be evaluated to give

$$E_G = 0.979\, W \text{ MeV} \tag{14}$$

with

$$W = A\, Z_0^2\, Z_1^2 \tag{14a}$$

$$A = A_0 A_1/(A_0 + A_1) \tag{14b}$$

in which A_0, A_1 are the atomic weights of the two colliding particles. For most nuclear reactions $S(E)$ is between 10 MeV-barns and 1 keV-barn.

The gas at a given r in the star has a given temperature so that the particles have a Boltzmann energy distribution. The rate of nuclear reactions is then proportional to

$$(8/\pi M)^{\frac{1}{2}} (k\,T)^{-3/2} \int \sigma(E)\, E \exp(-E/kT)\, \mathrm{d}E \tag{15}$$

It is most convenient[5] to write for the rate of disappearance of one of the reactants

$$\mathrm{d}X_0/\mathrm{d}t = -[01]\, X_0 X_1 \tag{16}$$

where X_0 and X_1 are the concentrations of the reactants by mass, and

$$[01] = 7.8 \cdot 10^{11} (Z_0 Z_1/A)^{1/3} S_{\text{eff}} \varrho T_6^{-2/3} e^{-\tau} \tag{17}$$

$$\tau = 42.487 (W/T_6)^{1/3} \tag{17a}$$

Since the reaction cross section (12) increases rapidly with energy, the main contribution to the reaction comes from particles which have an energy many times the average thermal energy. Indeed the most important energy is

$$E_0 = (\tau/3)\, k\, T \tag{18}$$

For $T = 13$ which is an average for the interior of the sun, we have

$$\begin{aligned} \tau/3 &= 4.7 \text{ for the reaction H+H} \\ &\quad 19 \text{ for the reaction C+H} \\ &\quad 25 \text{ for the reaction N+H} \end{aligned} \tag{19}$$

It is also easy to see from (17) that the temperature dependence of the reaction rate is

$$\frac{\mathrm{d} \ln [01]}{\mathrm{d} \ln T} = \frac{\tau - 2}{3} \tag{20}$$

Nuclear Reactions in Main Sequence Stars

Evidently, at a given temperature and under otherwise equal conditions, the reactions which can occur most easily are those which have the smallest possible value of W (14a). This means that at least one of the interacting nuclei should be a proton, $A_0 = Z_0 = 1$. Thus we may examine the reactions involving protons.

The simplest of all possible reactions is

$$H + H = D + \varepsilon^+ + \nu \tag{21}$$

(ε^+ = positron, ν = neutrino).

This was first suggested by von Weizsäcker[3], and calculated by Critchfield and Bethe[6]. The reaction is of course exceedingly slow because it involves the beta disintegration. Indeed the characteristic factor S is

$$S(E) = 3.36 \cdot 10^{-25} \text{ MeV-barns} \tag{22}$$

This has been derived on purely theoretical grounds, using the known coupling constant of beta disintegration; the value is believed to be accurate to 20% or better. There is no chance of observing such a slow reaction on earth, but

in the stars we have almost unlimited time, and a large supply of protons of high energy. As we shall see presently, the rate of energy production by this simple reaction fits the observed energy production in the sun very well.

The deuterons formed in (21) will quickly react further, and the end product is ^{4}He. We shall discuss the reactions in more detail later on.

The proton–proton reaction (21), although it predicts the correct energy production in the sun, has a rather weak dependence on temperature. According to (19), (20), it behaves about as T^4. Since central temperatures change only little from the sun to more massive stars, the energy production by this reaction does likewise. However as we have seen in Fig. 2, the observed energy production increases dramatically with increasing mass. Therefore there must exist nuclear reactions which are more strongly dependent on temperature; these must involve heavier nuclei.

Stimulated by the Washington Conference of April 1938, and following the argument just mentioned, I examined[7] the reactions between protons and other nuclei, going up in the periodic system. Reactions between H and ^{4}He lead nowhere, there being no stable nucleus of mass 5. Reactions of H with Li, Be and B, as well as with deuterons, are all very fast at the central temperature of the sun, but just this speed of the reaction rules them out: the partner of H is very quickly used up in the process. In fact, and just because of this reason, all the elements mentioned, from deuterium to boron, are extremely rare on earth and in the stars, and can therefore not be important sources of energy.

The next element, carbon, behaves quite differently. In the first place, it is an abundant element, probably making up about 1% by mass of any newly formed star. Secondly, in a gas of stellar temperature, it undergoes a cycle of reactions, as follows

$$^{12}\mathrm{C} + \mathrm{H} = {}^{13}\mathrm{N} + \gamma \qquad (23a)$$

$$^{13}\mathrm{N} = {}^{13}\mathrm{C} + \varepsilon^+ + \nu \qquad (23b)$$

$$^{13}\mathrm{C} + \mathrm{H} = {}^{14}\mathrm{N} + \gamma \qquad (23c)$$

$$^{14}\mathrm{N} + \mathrm{H} = {}^{15}\mathrm{O} + \gamma \qquad (23d)$$

$$^{15}\mathrm{O} = {}^{15}\mathrm{N} + \varepsilon^+ + \nu \qquad (23e)$$

$$^{15}\mathrm{N} + \mathrm{H} = {}^{12}\mathrm{C} + {}^4\mathrm{He} \qquad (23f)$$

Reactions a, c, and d are radiative captures; the proton is captured by the nucleus and the energy emitted in the form of gamma rays; these are then quickly converted into thermal energy of the gas. For reactions of this type, $S(E)$ is of the order of 1 keV-barn. Reactions b and e are simply spontaneous beta decays, with lifetimes of 10 and 2 min respectively, negligible in com-

parison with stellar times. Reaction f is the most common type of nuclear reaction, with 2 nuclei resulting from the collision; $S(E)$ for such reactions is commonly of the order of MeV-barns.

Reaction f is in a way the most interesting because it closes the cycle: we reproduce the ^{12}C which we started from. In other words, carbon is only used as a catalyst; the result of the reaction is a combination of 4 protons and 2 electrons[8] to form one ^{4}He nucleus. In this process two neutrinos are emitted, taking away about 2 MeV energy together. The rest of the energy, about 25 MeV per cycle, is released usefully to keep the sun warm.

Making reasonable assumptions of the reaction strength $S(E)$, on the basis of general nuclear physics, I found in 1938 that the carbon–nitrogen cycle gives about the correct energy production in the sun. Since it involves nuclei of relatively high charge, it has a strong temperature dependence, as given in (19). The reaction with ^{14}N is the slowest of the cycle and therefore determines the rate of energy production; it goes about as T^{24} near solar temperature. This is amply sufficient to explain the high rate of energy production in massive stars[9].

Experimental Results

To put the theory on a firm basis, it is important to determine the strength factor $S(E)$ for each reaction by experiment. This has been done under the leadership of W. A. Fowler[10] of the California Institute of Technology in a monumental series of papers extending over a quarter of a century. Not only have all the reactions in (23) been observed, but in all cases $S(E)$ has been accurately determined.

The main difficulty in this work is due to the resonances which commonly occur in nuclear reactions. Fig. 3 shows the cross section of the first reaction[5] (23a), as a function of energy. The measured cross sections extend over a factor of 10^7 in magnitude; the smallest ones are 10^{-11} barns $= 10^{-35}$ cm^2 and therefore clearly very difficult to observe. The curve shows a resonance at 460 keV. The solid curve is determined from nuclear reaction theory, on the basis of the existence of that resonance. The fit of the observed points to the calculated curve is impressive. Similar results have been obtained on the other three proton–capture reactions in (23).

On the basis of Fig. 3 we can confidently extrapolate the measurements to lower energy. As we mentioned in (18), the most important energy contribut-

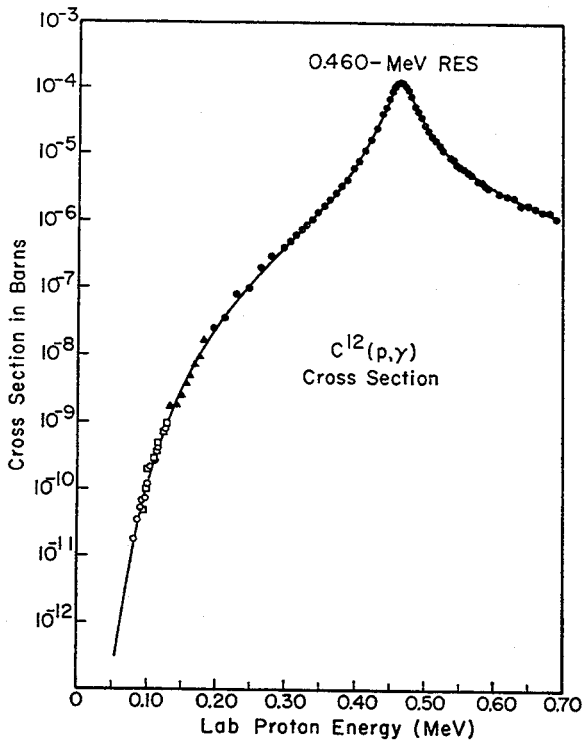


Fig. 3. Cross section for the reaction $^{12}C + H$, as a function of the proton energy. From Fowler, Caughlan and Zimmerman[5].

ing to the reaction rate is about 20 kT. For $T_6 = 13$, we have $kT = 1.1$ keV; so we are most interested in the cross section around 20 keV. This is much too low an energy to observe the cross section in the laboratory; even at 100 keV, the cross section is barely observable. So quite a long extrapolation is required. This can be done with confidence provided there are no resonances close to $E = 0$. Therefore a great deal of experimental work has gone into the search for such resonances.

The resonances exist of course in the compound nucleus, *i.e.* the nucleus obtained by adding the two initial reactants. To find resonances near the threshold of the reactions (23), it is necessary to produce the same compound nucleus from other initial nuclei, *e.g.*, in the reaction between ^{14}N and H, the compound nucleus ^{15}O is formed. To investigate its levels Hensley[11] at CalTech studied the reaction

$$^{16}O + {}^{3}He = {}^{15}O + {}^{4}He \qquad (24)$$

He found indeed a resonance 20 keV below the threshold for $^{14}N+H$ which in principle might enhance the process (23d). However the state in ^{15}O was found to have a spin $J=7/2$. Therefore, even though ^{14}H has $J=1$ and the proton has a spin of $1/2$, we need at least an orbital momentum $\lambda=2$, to reach this resonant state in ^{15}O. The cross section for such a high orbital momentum is reduced by at least a factor 10^4, compared to $\lambda=0$, so that the near-resonance does not in fact enhance the cross section $^{14}N+H$ appreciably. This cross section can then be calculated by theoretical extrapolation from the measured range of proton energies, and the same is true for the other reactions in the cycle (23).

On this basis, Fowler and others have calculated the rate of reactions in the CN cycle. A convenient tabulation has been given by Reeves[12]; his results are plotted in Fig. 4. This figure gives the energy production per gram per second as a function of temperature. We have assumed $X=0.5$, $Z=0.02$.

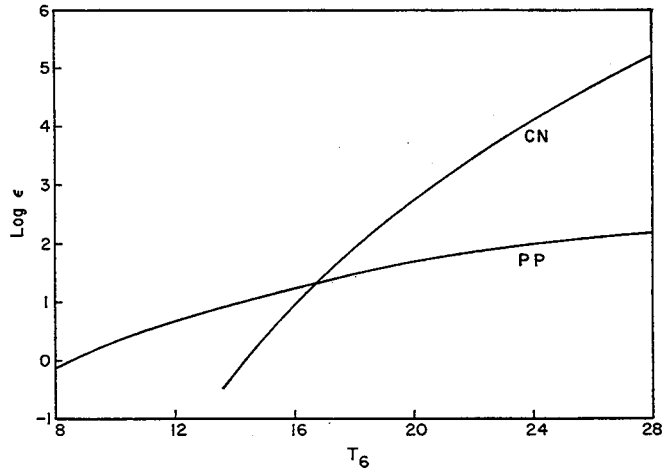


Fig. 4. The energy production, in erg/g · sec as a function of the temperature in millions of degrees. For the proton–proton reaction (PP) and the carbon–nitrogen cycle (CN). Concentrations assumed $X=Y=0.5$, $Z=0.02$. Calculated from Tables 8 and 9 of Reeves[12].

The figure shows that at low temperature the H+H reaction dominates, at high temperatures the C+N cycle; the crossing point is at $T_6=13$; here the energy production is 7 erg/g sec. The average over the entire sun is obviously smaller, and the result is compatible with an average production of 2 erg/g sec.

The energy production in the main sequence can thus be considered as well understood.

An additional point should be mentioned. Especially at higher temperature, when the CN cycle prevails, there is also a substantial probability for the reaction chain

$$^{16}\mathrm{O} + \mathrm{H} = {}^{17}\mathrm{F} + \gamma \tag{25a}$$

$$^{17}\mathrm{F} = {}^{17}\mathrm{O} + \varepsilon^{+} + \nu \tag{25b}$$

$$^{17}\mathrm{O} + \mathrm{H} = {}^{14}\mathrm{N} + {}^{4}\mathrm{He} \tag{25c}$$

This chain is not cyclic but feeds into the CN cycle. It is customary to speak of the whole set of reactions as the CNO bi-cycle. The effect of reactions (25) is that ^{16}O initially present will also contribute to the reactants available, and thus increase the reaction rate of the CN cycle somewhat. This has been taken into account in Fig. 5.

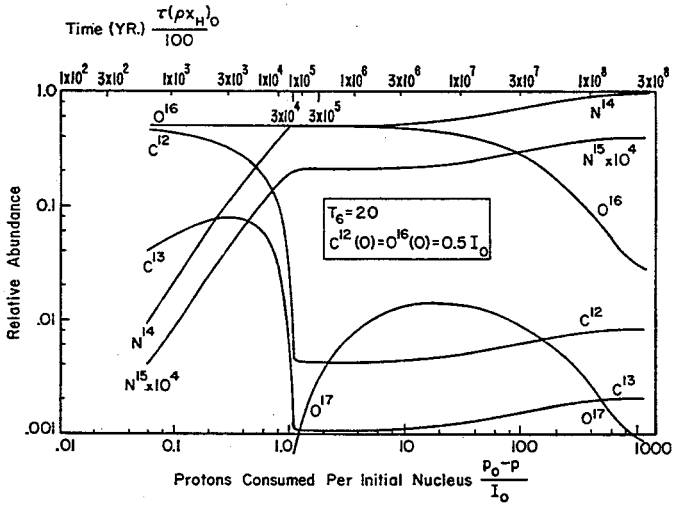


Fig. 5. Variation with time of the abundances of various elements involved in the CNO cycle. It is assumed that initially ^{12}C and ^{16}O have the same abundance while that of ^{14}N is small. From G.R. Caughlan, *Astrophys. J.* (1967).

If equilibrium is established in the CNO bi-cycle, eventually most of the nuclei involved will end up as ^{14}N because this nucleus has by far the longest lifetime against nuclear reactions. There is no observable evidence for this; in fact wherever the abundance can be observed, C and O tend to be at least

as abundant as N. However this is probably due to the fact that the interior of a star stays well separated from its surface; there is very little mixing. Astrophysicists have investigated the circumstances when mixing is to be expected, and have found that surface abundances are quite compatible with these expectations. In the interstellar material which is used to form stars, we have reason to believe that C and O are abundant and N is rare. This will be discussed later.

The Completion of the Proton-Proton Chain

The initial reaction (21) is followed almost immediately by

$$^2\mathrm{H} + \mathrm{H} = {}^3\mathrm{He} + \gamma \tag{26}$$

The fate of ^{3}He depends on the temperature. Below about $T_6 = 15$, the ^{3}He builds up sufficiently so that such nuclei react with each other according to

$$2\ {}^3\mathrm{He} = {}^4\mathrm{He} + 2\ \mathrm{H} \tag{27}$$

This reaction has an unusually high $S(E) = 5$ MeV-barns[5]. At higher temperature, the reaction

$$^4\mathrm{He} + {}^3\mathrm{He} = {}^7\mathrm{Be} + \gamma \tag{28}$$

competes favorably with (27). The ^{7}Be thus formed may again react in one of two ways

$$^7\mathrm{Be} + e^- = {}^7\mathrm{Li} + \nu \tag{29a}$$

$$^7\mathrm{Be} + \mathrm{H} = {}^8\mathrm{B} + \gamma \tag{29b}$$

$$^8\mathrm{B} = 2\ {}^4\mathrm{He} + \varepsilon^+ + \nu \tag{29c}$$

At about $T_C = 20$, reaction (29b) begins to dominate over (29a). (29b) is followed by (29c) which emits neutrinos of very high energy. Davies[13], at Brookhaven, is attempting to observe these neutrinos.

Evolution of a Star

A main sequence star uses up its hydrogen preferentially near its center where nuclear reactions proceed most rapidly. After a while, the center has lost al-

most all its hydrogen. For stars of about twice the luminosity of the sun, this happens in less than 10^{10} years which is approximately the age of the universe, and also the age of stars in the globular clusters. We shall now discuss what happens to a star after it has used up the hydrogen at the center. Of course, in the outside regions hydrogen is still abundant.

This evolution of a star was first calculated by Schwarzschild[14] who has been followed by many others; we shall use recent calculations by Iben[15]. When hydrogen gets depleted, not enough energy is produced near the center to sustain the pressure of the outside layers of the star. Hence gravitation will cause the center to collapse. Thereby, higher temperatures and densities are achieved. The temperature also increases farther out where there is still hydrogen left, and this region now begins to burn. After a relatively short time, a shell of *H*, away from the center, produces most of the energy; this shell gradually moves outward and gets progressively thinner as time goes on.

At the same time, the region of the star outside the burning shell expands. This result follows clearly from all the many numerical computations on this subject. The physical reason is not clear. One hypothesis is that it is due to the discontinuity in mean molecular weight: Inside the shell, there is mostly helium, of $\mu = 4/3$, outside we have mostly hydrogen, and $\mu = 0.65$. Another suggestion is that the flow of radiation is made difficult by the small radius

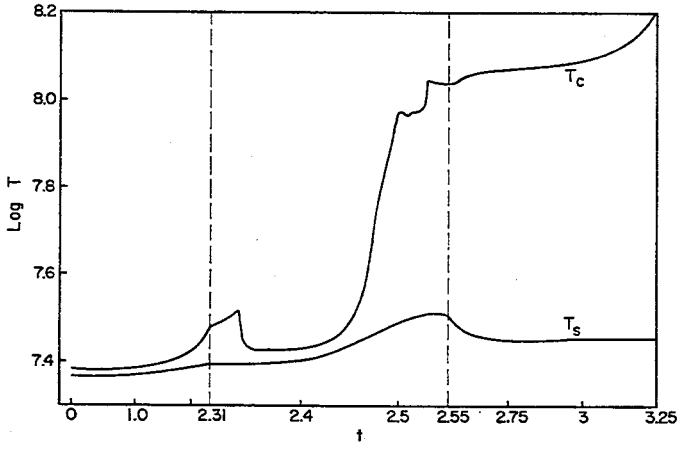


Fig. 6. Evolution of a star of 3 solar masses, according to I. Iben, *Astrophys. J.*, 142 (1965) 1447. Abscissa is time in units of 10^8 years (note the breaks in scale at $t = 2.31$ and 2.55). I. Temperature (on logarithmic scale): T_c = temperature at center of star, T_s = same at mid-point of source of energy generation, which, after $t = 2.48$ is a thin shell. T_c increases enormously, T_s stays almost constant.

of the energy source, and that this has to be compensated by lower density ust outside the source.

By this expansion the star develops into a red giant. Indeed, in globular clusters (which, as I mentioned, are made up of very old stars), all the more luminous stars are red giants. In the outer portion of these stars, radiative transport is no longer sufficient to carry the energy flow; therefore convection of material sets in in these outer regions. This convection can occupy as much as the outer 80% of the mass of the star; it leads to intimate mixing of the material in the convection zone.

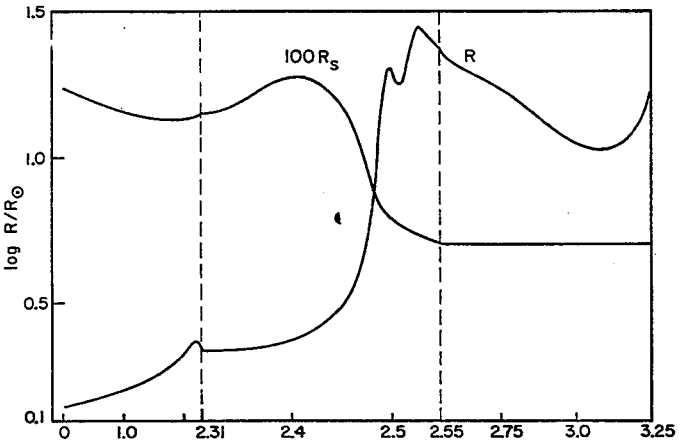


Fig. 7. Evolution of a star, (see caption to Fig. 6). II Radius, in units of that of the sun, on logarithmic scale. R = total radius, 100 R_S = 100 times the radius of mid-point of energy source. R increases tremendously, while R_S shrinks somewhat.

Iben[15] has discussed a nice observational confirmation of this convectional mixing. The star Capella is a double star, each component having a mass of about 3 solar masses, and each being a red giant. The somewhat lighter star, «Capella F» (its spectral type is F) shows noticeable amounts of Li in its spectrum, while the somewhat heavier Capella G shows at least 100 times less Li. It should be expected that G, being heavier, is farther advanced in its evolution. Iben now gives arguments that the deep-reaching convection and mixing which we just discussed, will occur just between the evolution phases F and G. By convection, material from the interior of the star will be carried to the surface; this material has been very hot and has therefore burned up its Li. Before deep convection sets in (in star F) the surface Li never sees high temperature and thus is preserved.

Following the calculations of Iben we have plotted in Figs. 6–9 the development of various important quantities in the history of a star of mass = 3 solar masses. The time is in units of 10^8 years. Since the developments go at very variable speed, the time scale has been broken twice, at $t = 2.31$ and $t = 2.55$. In between is the period during which the shell source develops.

During this period the central temperature rises spectacularly (Fig. 6), from about $T_6 = 25$ to $T_6 = 100$. At the same time the radius increases from about 2 to 30 solar radii; subsequently, it decreases again to about 15 (Fig. 7). The central density, starting at about 40, increases in the same period to about $5 \cdot 10^4$ (Fig. 8). The luminosity (Fig. 9) does not change spectacularly, staying always between 100 and 300 times that of the sun.

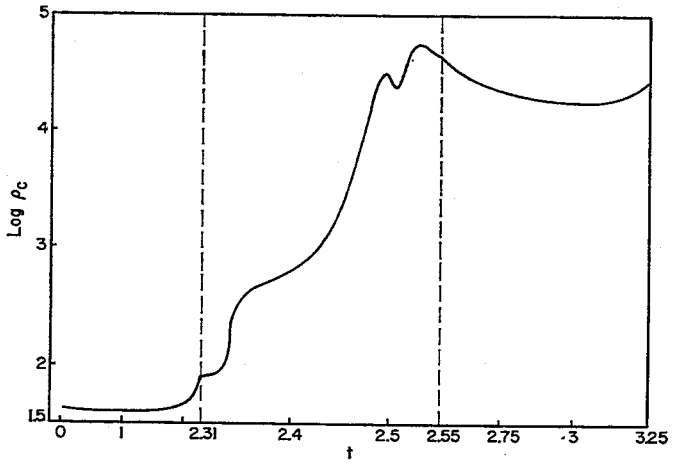


Fig. 8. Evolution of a star (see caption to Fig. 6).III. Density, on logarithmic scale, at the center of the star. This quantity increases about 1000-fold.

While the inside and the outside of the star undergo such spectacular changes, the shell in which the hydrogen is actually burning, does not change very much. Fig. 9 shows m, the fraction of the mass of the star enclosed by the burning shell. Even at the end of the calculation, $t = 3.25$, this is only $m = 0.2$. This means that only 20% of the hydrogen in the star has burned after all this development. Fig. 6, curve T_s, shows the temperature in the burning shell which stays near 25 million degrees all the time. Fig. 7, curve R_s, shows the radius of the shell, in units of the solar radius; during the critical time when the shell is formed this radius drops from about 0.15 to 0.07. This is of course the mechanism by which the shell is kept at the temperature which originally prevailed at the center.

In the meantime, the temperature at the center increases steadily. When it reaches about $T_6 = 100$, the ^{4}He which is abundant at the center, can undergo nuclear reactions. The first of these, which occurs at the lowest temperature (about $T_6 = 90$) is

$$^{14}\mathrm{N} + {}^4\mathrm{He} = {}^{18}\mathrm{F} + \gamma \qquad (30)$$

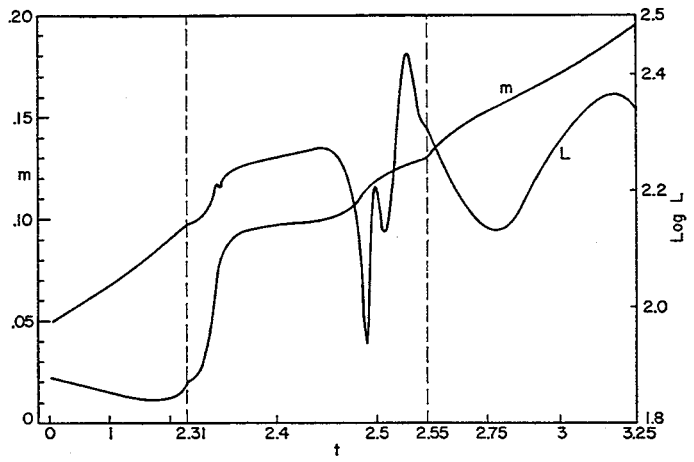


Fig. 9. Evolution of a star (see caption to Fig. 6). IV. Curve L, luminosity relative to that of the sun, on logarithmic scale. This quantity does not change very much during the life of the star. Curve m, fraction of the mass of the star enclosed by energy-producing shell, on linear scale. This fraction increases slowly with time.

While this reaction goes on, the central temperature remains fairly constant. However, there is not much ^{14}N so the reaction soon stops (after about $0.02 \cdot 10^8$ years), and the center contracts further.

The next reaction makes use entirely of the abundant ^{4}He, *viz.*

$$3\ {}^4\mathrm{He} = {}^{12}\mathrm{C} + \gamma \qquad (31)$$

This reaction has the handicap of requiring a simultaneous collision of 3 alpha particles. This would be extremely unlikely were it not for the fact that it is favored by a *double* resonance. Two alpha particles have nearly the same energy as the unstable nucleus ^{8}Be, and further $^8\mathrm{Be} + {}^4\mathrm{He}$ has almost the same energy as an excited state of ^{12}C. This reaction can of course not be observed in the laboratory but the two contributing resonances can be. The importance of the first resonance was first suggested by Salpeter[16], the second by Hoyle[17]. Recent data indicate that (31) requires a temperature of about $T_6 = 110$, at the

central densities corresponding to $t = 2.5$, *i.e.* $\varrho_c > 10^4$. Once this reaction sets in, the central temperature does not rise very fast any more.

Reaction (31) is most important for the buildup of elements. Early investigators[3,7] had great trouble with bridging the gap between ^{4}He and ^{12}C. Two nuclei in this gap, mass 5 and mass 8, are completely unstable, the rest disintegrate in a very short time under stellar conditions. Reaction (31) however leads to stable ^{12}C. This nucleus can now capture a further alpha particle

$$^{12}\mathrm{C} + {}^4\mathrm{He} = {}^{16}\mathrm{O} + \gamma \qquad (32)$$

the temperatures required for this are about the same as for (31). There is also some capture of alpha particles by ^{16}O leading to ^{20}Ne, but the next step, $^{20}\mathrm{Ne} \rightarrow {}^{24}\mathrm{Mg}$, cannot occur appreciably at these temperatures; instead, the helium gets used up in forming ^{12}C, ^{16}O and some ^{20}Ne.

Helium is depleted first in the center, and now the same process repeats which previously took place with hydrogen. A shell of burning He is formed, at a smaller radius than the H shell, and of course at a higher temperature. The center of the star now contracts further by gravitation and reaches still higher temperatures.

Buildup and Dispersal of Elements

The further developments of a massive star are more speculative. However the theory of Hoyle and collaborators[18] is likely to be correct.

The center of the star heats up until the newly formed carbon nuclei can react with each other. This happens at a temperature of roughly 10^9 degrees. Nuclei like ^{24}Mg or ^{28}Si can be formed. There are also more complicated mechanisms in which we first have a capture reaction with emission of a gamma ray, followed by capture of this gamma ray in another nucleus which releases ^{4}He. This ^{4}He can then enter further nuclei and build up the entire chain of stable nuclei up to the most stable Fe. Not much energy is released in all of these processes.

The center of the star contracts further and gets still hotter. At very high temperatures, several milliards of degrees, thermal equilibrium is no longer strongly in favor of nuclei of the greatest binding energy. Instead, endothermic processes can take place which destroy some of the stable nuclei already formed. In the process, alpha particles, protons and even neutrons may be released. This permits the buildup of elements beyond Fe, up to the top of the

periodic table. Because of the high temperatures involved all this probably goes fairly fast, perhaps in thousands of years.

During this stage, nuclear processes tend to consume rather than release energy. Therefore they no longer oppose the gravitational contraction so that contraction continues unchecked. It is believed that this will lead to an unstable situation. Just as the first contraction, at the formation of the H shell source, led to an expansion of the outer envelope of the star, a similar outward expansion is expected now. But time scales are now short, and this expansion may easily be an explosion. Hoyle *et al.*[18] have suggested this as the mechanism for a supernova.

In a supernova explosion much of the material of the star is ejected into interstellar space. We see this, *e.g.*, in the Crab Nebula. The ejected material probably contains the heavy elements which have been formed in the interior of the massive star. Thus heavy elements get into the interstellar gas, and can then be collected again by newly forming stars. It is believed that this is the way how stars get their heavy elements. This means that most of the stars we see, including our sun, are at least second generation stars, which have collected the debris of earlier stars which have suffered a supernova explosion.

To clinch this argument it must be shown that heavy elements cannot be produced in other ways. This has indeed been shown by Fowler[19]. He has investigated the behavior of the enormous gas cloud involved in the original «Big Bang», and its development with time. He has shown that temperatures and densities, as functions of time, are such that heavy elements beginning with C cannot be produced. The only element which can be produced in the big bang is ^{4}He.

If all this is true, stars have a life cycle much like animals. They get born, they grow, they go through a definite internal development, and finally they die, to give back the material of which they are made so that new stars may live.

I am very grateful to Professor E. E. Salpeter for his extensive help in preparing this paper.

1. R. d'E. Atkinson and F. G. Houtermans, *Z. Physik*, 54 (1929) 656.
2. G. Gamow and E. Teller, *Phys. Rev.*, 53 (1938) 608.
3. C. F. von Weizsäcker, *Physik. Z.*, 38 (1937) 176.

4. P. R. Demarque and J. R. Percy, *Astrophys. J.*, 140 (1964) 541.
5. W. A. Fowler, G. R. Caughlan and B. A. Zimmerman, *Ann. Rev. Astron. Astrophys.*, 5 (1967) 525.
6. H. A. Bethe and C. L. Critchfield, *Phys. Rev.*, 54 (1938) 248.
7. H. A. Bethe, *Phys. Rev.*, 55 (1939) 436.
8. The electrons are used to annihilate the positrons emitted in reactions b and e.
9. The carbon–nitrogen cycle was also discovered independently by C. F. von Weizsäcker, *Physik. Z.*, 39 (1938) 633, who recognized that this cycle consumes only the most abundant element, hydrogen. But he did not investigate the rate of energy production or its temperature dependence.
10. W. A. Fowler, many papers in *Phys. Rev.*, *Astrophys. J.* and other publications. Some of this work is summarized in ref. 5.
11. D. C. Hensley, *Astrophys. J.*, 147 (1967) 818.
12. H. Reeves, in Stellar Structure, Vol. 8 of G. P. Kuiper (Ed.), *Stars and Stellar Systems*, University of Chicago Press, Chicago, Ill., 1965, especially Tables 8 and 9.
13. R. Davies Jr., *Phys. Rev. Letters*, 12 (1964) 303.
14. M. Schwarzschild, *Structure and Evolution of the Stars*, Princeton University Press, 1958.
15. I. Iben Jr., *Astrophys. J.*, 141 (1965) 993, 142 (1965) 1447, 143 (1966) 483.
16. E. E. Salpeter, *Phys. Rev.*, 88 (1952) 547.
17. F. Hoyle, *Astrophys. J.*, Suppl. 1 (1954) 121.
18. E. M. Burbidge, G. R. Burbidge, W. A. Fowler and F. Hoyle, *Rev. Mod. Phys.*, 29 (1957) 547.
19. W. A. Fowler, *Internat. Ass. Geochem. Cosmochem., 1st Meeting, Paris, 1967.*

Biography

235

Hans Albrecht Bethe was born in Strasbourg, Alsace-Lorraine, on July 2 1906. He attended the Gymnasium in Frankfurt from 1915 to 1924. He then studied at the University of Frankfurt for two years, and at Munich for two and one-half years, taking his Ph. D. in theoretical physics with Professor Arnold Sommerfeld in July 1928.

He then was an Instructor in physics at Frankfurt and at Stuttgart for one semester each. From fall 1929 to fall 1933 his headquarters were the University of Munich where he became Privatdozent in May 1930. During this time he had a travel fellowship of the International Education Board to go to Cambridge, England, in the fall of 1930, and to Rome in the spring terms of 1931 and 1932. In the winter semester of 1932–1933, he held a position as Acting Assistant Professor at the University of Tübingen which he lost due to the advent of the Nazi regime in Germany.

Bethe emigrated to England in October 1933 where he held a temporary position as Lecturer at the University of Manchester for the year 1933–1934, and a fellowship at the University of Bristol in the fall of 1934. In February 1935 he was appointed Assistant Professor at Cornell University, Ithaca, N. Y. U. S. A., then promoted to Professor in the summer of 1937. He has stayed there ever since, except for sabbatical leaves and for an absence during World War II. His war work took him first to the Radiation Laboratory at the Massachusetts Institute of Technology, working on microwave radar, and then to the Los Alamos Scientific Laboratory which was engaged in assembling the first atomic bomb. He returned to Los Alamos for half a year in 1952. Two of his sabbatical leaves were spent at Columbia University, one at the University of Cambridge, and one at CERN and Copenhagen.

Bethe's main work is concerned with the theory of atomic nuclei. Together with Peierls, he developed a theory of the deuteron in 1934 which he extended in 1949. He resolved some contradictions in the nuclear mass scale in 1935. He studied the theory of nuclear reactions in 1935–1938, predicting many reaction cross sections. In connection with this work, he developed Bohr's theory of the compound nucleus in a more quantitative fashion. This work

and also the existing knowledge on nuclear theory and experimental results, was summarized in three articles in the *Reviews of Modern Physics* which for many years served as a textbook for nuclear physicists.

His work on nuclear reactions led Bethe to the discovery of the reactions which supply the energy in the stars. The most important nuclear reaction in the brilliant stars is the carbon–nitrogen cycle, while the sun and fainter stars use mostly the proton–proton reaction. Bethe's main achievement in this connection was the exclusion of other possible nuclear reactions. The Nobel Prize was given for this work, as well as his work on nuclear reactions in general.

In 1955 Bethe returned to the theory of nuclei, emphasizing a different phase. He has worked since then on the theory of nuclear matter whose aim it is to explain the properties of atomic nuclei in terms of the forces acting between nucleons.

Before his work on nuclear physics, Bethe's main attention was given to atomic physics and collision theory. On the former subject, he wrote a review article in *Handbuch der Physik* in which he filled in the gaps of the existing knowledge, and which is still up-to-date. In collision theory, he developed a simple and powerful theory of inelastic collisons between fast particles and atoms which he has used to determine the stopping power of matter for fast charged particles, thus providing a tool to nuclear physicists. Turning to more energetic collisions, he calculated with Heitler the Bremsstrahlung emitted by relativistic electrons, and the production of electron pairs by high energy gamma rays.

Bethe also did some work on solid-state theory. He discussed the splitting of atomic energy levels when an atom is inserted into a crystal, he did some work on the theory of metals, and especially he developed a theory of the order and disorder in alloys.

In 1947, Bethe was the first to explain the Lamb-shift in the hydrogen spectrum, and he thus laid the foundation for the modern development of quantum electrodynamics. Later on, he worked with a large number of collaborators on the scattering of pi mesons and on their production by electromagnetic radiation.

Bethe is married to the daughter of P. P. Ewald, the well-known X-ray physicist. They have two children, Henry and Monica.

Physics 1968

237

LUIS W. ALVAREZ

«for his decisive contributions to elementary particle physics, in particular the discovery of a large number of resonance states, made possible through his development of the technique of using hydrogen bubble chamber and data analysis»

Physics 1968

Presentation Speech by Professor S. von Friesen, member of the Swedish Academy of Sciences

Your Majesty, Your Royal Highnesses, Ladies and Gentlemen.

The science of physics has as its function the study of energy in all its forms. Einstein observed that matter, or mass, is one of the forms in which energy manifests itself. This fact was established experimentally 35 years ago, when it was discovered that high-energy electromagnetic radiation was capable of producing pairs of electrons, one with positive, the other with negative charges. It has since been possible to produce other similar pairs, for example protons and antiprotons. These newly-created particles are stable and, if left undisturbed, can exist indefinitely. Unstable particles can also be produced, however. These disintegrate rapidly into other particles and, passing through one or several stages, revert to stable forms or develop into other forms of energy. Many such new particles have been discovered and studied during the last two decades. They are so minute that it is impossible to see them; they can only be identified by the tracks they leave behind them as they move. The scientist must behave like the hunter, who determines the identity and behaviour of his quarry by studying tracks left in the snow.

The new particles are normally produced with the help of the great, new accelerators which cause the particles to move at very great speed. This has the advantage that, although the life-span of the particle might be as little as a ten-thousandth part of a millionth of a second, the track acquires a length of several centimetres.

One could, however, suspect the existence of particles with considerably shorter life-spans and with such small track-lengths that they are impossible to measure. In this case one is obliged, instead, to study the tracks of their disintegration products and the tracks of the reactions they produce in collision with other particles. The pattern of tracks thus becomes very complicated; the correct interpretation of what actually occurs requires acute powers of discernment and a particularly advanced experimental technique. It is in this field that Professor Luis Alvarez has made the contributions for which he is today being rewarded.

He has with insight and determination developed the bubble-chamber,

invented by the Nobel Prize winner in Physics, Donald Glaser, into an invaluable instrument for this type of investigation. Alvarez' bubble-chamber contains many hundreds of litres of hydrogen, reduced to a temperature of minus 250 °C, which thus becomes fluid. When the particle passes through the liquid, it is warmed to boiling point along the track it leaves. In the wake are a trail of bubbles that can be photographed whilst still very small. The photographs are able, in this way, to reproduce accurately the path of the particle. Because the chamber contains only hydrogen, it is evident that all reactions must occur with hydrogen nuclei, protons. This considerably simplifies the interpretation of the phenomenon. The cost of this instrument, capable of producing about a million photographs annually, was two million dollars.

The photographs must be studied and measured with great accuracy. In order to carry out this enormous task, Alvarez and his assistants have constructed a series of more and more delicate automatic scanning and measuring instruments capable of transferring the information from the photographic film into a state suitable for treatment by computer. In this field, too, Alvarez has made contributions of a pioneering nature.

With the establishment of the hydrogen bubble-chamber, entirely new possibilities for research into high-energy physics present themselves. Results have already been apparent in the form of newly-discovered elementary particles. The first, very short-lived, so called, «resonance particle» was found in 1960. Since then there have been a whole series of discoveries made by Alvarez' group in Berkeley, California and in other laboratories where Alvarez' material is being used or where his methods and programs are adopted. Practically all the discoveries that have been made in this important field of high-energy physics have been possible only through the use of methods originated by Professor Alvarez.

Dr. Alvarez. Your contributions to physics are numerous and important. To-day our attention is focused on the outstanding discoveries which you have made in the field of high-energy physics as a result of your far-sighted and bold development of the hydrogen bubble-chamber into an instrument of great power and high precision and of the means of handling and analysing the large quantities of valuable information which it can produce.

On behalf of the Royal Swedish Academy of Sciences I extend to you our warm congratulations and now ask you to receive the Nobel Prize from the hands of His Majesty the King.

LUIS W. ALVAREZ

Recent developments in particle physics

Nobel Lecture, December 11, 1968

When I received my B. S. degree in 1932, only two of the fundamental particles of physics were known. Every bit of matter in the universe was thought to consist solely of protons and electrons. But in that same year, the number of particles was suddenly doubled. In two beautiful experiments, Chadwick[1] showed that the neutron existed, and Anderson[2] photographed the first unmistakable positron track. In the years since 1932, the list of known particles has increased rapidly, but not steadily. The growth has instead been concentrated into a series of spurts of activity.

Following the traditions of this occasion, my task this afternoon is to describe the latest of these periods of discovery, and to tell you of the development of the tools and techniques that made it possible. Most of us who become experimental physicists do so for two reasons; we love the tools of physics because to us they have intrinsic beauty, and we dream of finding new secrets of nature as important and as exciting as those uncovered by our scientific heroes. But we walk a narrow path with pitfalls on either side. If we spend all our time developing equipment, we risk the appellation of «plumber», and if we merely use the tools developed by others, we risk the censure of our peers for being parasitic. For these reasons, my colleagues and I are grateful to the Royal Swedish Academy of Science for citing both aspects of our work at the Lawrence Radiation Laboratory at the University of California – the observations of a new group of particles and the creation of the means for making those observations.

As a personal opinion, I would suggest that modern particle physics started in the last days of World War II, when a group of young Italians, Conversi, Pancini, and Piccioni, who were hiding from the German occupying forces, initiated a remarkable experiment. In 1946, they showed[3] that the «mesotron» which had been discovered in 1937 by Neddermeyer and Anderson[4] and by Street and Stevenson[5], was not the particle predicted by Yukawa[6] as the mediator of nuclear forces, but was instead almost completely unreactive in a nuclear sense. Most nuclear physicists had spent the war years in military-

related activities, secure in the belief that the Yukawa meson was available for study as soon as hostilities ceased. But they were wrong.

The physics community had to endure less than a year of this nightmarish state; Powell and his collaborators[7] discovered in 1947 a singly charged particle (now known as the pion) that fulfilled the Yukawa prediction, and that decayed into the «mesotron», now known as the muon. Sanity was restored to particle physics, and the pion was found to be copiously produced in Ernest Lawrence's 184-inch cyclotron, by Gardner and Lattes[8] in 1948. The cosmic ray studies of Powell's group were made possible by the elegant nuclear-emulsion technique they developed in collaboration with the Ilford laboratories under the direction of C. Waller.

In 1950, the pion family was filled out with its neutral component by three independent experiments. In Berkeley, at the 184-inch cyclotron, Moyer, York *et al.*[9] measured a Doppler-shifted γ-ray spectrum that could only be explained as arising from the decay of a neutral pion, and Steinberger, Panofsky and Steller[10] made the case for this particle even more convincing by a beautiful experiment using McMillan's new 300-MeV synchrotron. And independently at Bristol, Ekspong, Hooper and King[11] observed the two-γ-ray decay of the π^0 in nuclear emulsion, and showed that its lifetime was less than $5 \cdot 10^{-14}$ sec.

In 1952 Anderson, Fermi and their collaborators[12] at Chicago started their classic experiments on the pion–nucleon interaction at what we would now call low energy. They used the external pion beams from the Chicago synchrocyclotron as a source of particles, and discovered what was for a long time called *the* pion–nucleon resonance. The isotopic spin formalism, which had been discussed for years by theorists since its enunciation in 1936 by Cassen and Condon[13], suddenly struck a responsive chord in the experimental physics community. They were impressed by the way Brueckner[14] showed that «*I*-spin» invariance could explain certain ratios of reaction cross sections, if the resonance, which had been predicted many years earlier by Pauli and Dancoff[15], were in the 3/2 isotopic spin state, and had an angular momentum of 3/2.

By any test we can now apply, the «3,3-resonance» of Anderson, Fermi *et al.* was the first of the «new particles» to be discovered. But since the rules for determining what constitutes a discovery in physics have never been codified–as they have been in patent law–it is probably fair to say that it was not customary, in the days when the properties of the 3,3-resonance were of paramount importance to the high-energy physics community, to regard that

resonance as a «particle». Neutron spectroscopists study hundreds of resonances in neutron–nucleus systems which they do not regard as separate entities, even though their lives are billions of times as long. I don't believe that an early and general recognition that the 3,3-resonance should be listed in the «table of particles» would in any way have speeded up the development of high-energy physics.

Although the study of the production and the interaction of pions had passed in a decisive way from the cosmic-ray groups to the accelerator laboratories in the late 1940's, the cosmic-ray-oriented physicsts soon found two new families of «strange particles» – the *K* mesons and the hyperons. The existence of the strange particles has had an enormous impact on the work done by our group at Berkeley. It is ironic that the parameters of the Bevatron were fixed and the decision to build that accelerator had been made before a single physicist in Berkeley really believed in the existence of strange particles. But as we look back on the evidence, it is obvious that the observations were well made, and the conclusions were properly drawn. Even if we had accepted the existence – and more pertinently the importance – of these particles, we would not have known what energy the Bevatron needed to produce strange particles; the associated production mechanism of Pais[16] and its experimental proof by Fowler, Shutt *et al.*[17] were still in the future. So the fact that, with a few notable exceptions, the Bevatron has made its greatest contributions to physics in the field of strange particles must be attributed to a very fortunate set of accidents.

The Bevatron's proton energy of 6.3 GeV was chosen so that it would be able to produce antiprotons, if such particles could be produced. Since, in the interest of keeping the «list of particles» tractable, we no longer count antiparticles nor individual members of *I*-spin multiplets, it is becoming fashionable to regard the discovery of the antiproton as an «obvious exercise for the student». (If we were to apply the «new rules» to the classical work of Chadwick and Anderson, we would conclude that they hadn't done anything either – the neutron is simply another *I*-spin state of the proton, and Anderson's positron is simply the obvious antielectron!) In support of the non-obvious nature of the Segrè group's discovery of the antiproton[18] I need only recall that one of the most distinguished high-energy physicists I know, who didn't believe that antiprotons could be produced, was obliged to settle a 500-dollar bet with a colleague who held the now universally accepted belief that all particles can exist in an antistate.

I have just discussed in a very brief way the discovery of some particles that

have been of importance in our bubble-chamber studies, and I will continue the discussion throughout my lecture. This account should not be taken to be authoritative–there *is* no authority in this area–but simply as a narrative to indicate the impact that certain experimental work had on my own thinking and on that of my colleagues.

I will now return to the story of the very important strange particles. In contrast to the discovery of the pion, which was accepted immediately by almost everyone–one apparent exception will be related later in this talk–the discovery and the eventual acceptance of the existence of the strange particles stretched out over a period of a few years. Heavy, unstable particles were first seen in 1947, by Rochester and Butler[19],who photographed and properly interpreted the first two «*V* particles» in a cosmic-ray-triggered cloud chamber. One of the *V*'s was charged, and was probably a *K* meson. The other was neutral, and was probably a K^0. For having made these observations, Rochester and Butler are generally credited with the discovery of strange particles. There was a disturbing period of two years in which Rochester and Butler operated their chamber and no more *V* particles were found. But in 1950 Anderson, Leighton *et al.*[20] took a cloud chamber to a mountain top and showed that it was possible to observe approximately one *V* particle per day under such conditions. They reported, «To interpret these photographs, one must come to the same remarkable conclusion as that drawn by Rochester and Butler on the basis of these two photographs, *viz.*, that these two types of events represent, respectively, the spontaneous decay of neutral and charged unstable particles of a new type.»

Butler and his collaborators then took their chamber to the Pic-du-Midi and confirmed the high event rate seen by the CalTech group on White Mountain. In 1952 they reported the first cascade decay[21]–now known as the Ξ^- hyperon.

While the cloud-chamber physicists were slowly making progress in understanding the strange particles, a parallel effort was under way in the nuclear emulsion-oriented laboratories. Although the first *K* meson was undoubtedly observed in Leprince–Ringuet's cloud chamber [22] in 1944, Bethe[23] cast sufficient doubt on its authenticity that it had no influence on the physics community and on the work that followed. The first overpowering evidence for a *K* meson appeared in nuclear emulsion, in an experiment by Brown and most of the Bristol group[24], in 1949. This so-called τ^+ meson decayed at rest into three coplanar pions. The measured ranges of the three pions gave a very accurate mass value for the τ meson of 493.6 MeV. Again there was a disturb-

inert, nuclearly. The suggestion by Marshak and Bethe[29] that it was the daughter of a strongly interacting particle was published almost simultaneously with the independent experimental demonstration by Powell *et al.*[7] mentioned earlier. Although invoking a similar mechanism to bring order into the strange-particle arena was tempting, Pais[16] made his suggestion that strange particles were produced «strongly» in pairs, but decayed «weakly» when separated from each other.

Gell-Mann[30] (and independently Nishijima[31]) then made the first of his several major contributions to particle physics by correctly guessing the rules that govern the production and decay of all the strange particles. I use the word «guessing» with the same sense of awe I feel when I say that Champollion guessed the meanings of the hieroglyphs on the Rosetta Stone. Gell-Mann had first to assume that the K meson was not an I-spin triplet, as it certainly appeared to be, but an I-spin doublet plus its antiparticles, and he had further to assume the existence of the neutral Σ and of the neutral Ξ. And finally, when he assigned appropriate values of his new quantum number, strangeness, to each family, his rules explained the one observed production reaction and predicted a score of others. And of course it explained all the known decays, and predicted another. My research group eventually confirmed all of Gell-Mann's and Nishijima's early predictions, many of them for the first time, and we continue to be impressed by their simple elegance.

This was the state of the art in particle physics in 1954, when William Brobeck turned his brainchild, the Bevatron, over to his Radiation Laboratory associates to use as a source of high-energy protons. I had been using the Berkeley proton linear accelerator in some studies of short-lived radioactive species, and I was pleased at the chance to switch to a field that appeared to be more interesting. My first Bevatron experiment was done in collaboration with Sula Goldhaber[32]; it gave the first real measurement of the τ meson lifetime. My next experiment was done with three talented young postdoctoral fellows, Frank S. Crawford Jr., Myron L. Good and M. Lynn Stevenson. An early puzzle in K-meson physics was that two of the particles (the θ and τ) had similar, but poorly determined, lifetimes and masses. That story has been told in his auditorium by Lee[33] and Yang[34], so I won't repeat it now. But I do like to think that our demonstration[35], simultaneously with and independently from one by Fitch and Motley[36], that the two lifetimes were not measurably different, plus similar small limits on possible mass differences set by von Friesen *et al.*[37] and by Birge *et al.*[38], nudged Lee and Yang a bit toward their revolutionary conclusion.

Our experiences with what was then a very complicated array of scintillation counters led me and my colleagues to despair of making meaningful measurements of what we perceived to be the basic reactions of strange particle physics:

$$\begin{array}{ccccc} \pi^- + p \rightarrow & \Lambda & + & K^0 \\ & \downarrow & & \downarrow \\ & p + \pi^- & & \pi^- + \pi^+ \end{array}$$

The production reaction is indicated by the horizontal arrows, the subsequent decays by the vertical arrows. Fig.1 shows a typical example of this reaction, as we saw it later in the 10-inch bubble chamber. We concluded, correctly I believe, that none of the then known techniques was well suited to study this reaction. Counters appeared hopelessly inadequate to the task, and the spark chamber had not yet been invented. The Brookhaven diffusion-cloud-chamber group[17] had photographed only a few events like that shown in Fig.1, in a period of two years. It seemed to us that a track-recording technique was called for, but each of the three known track devices had drawbacks that ruled it out as a serious contender for the role we envisaged. Nuclear

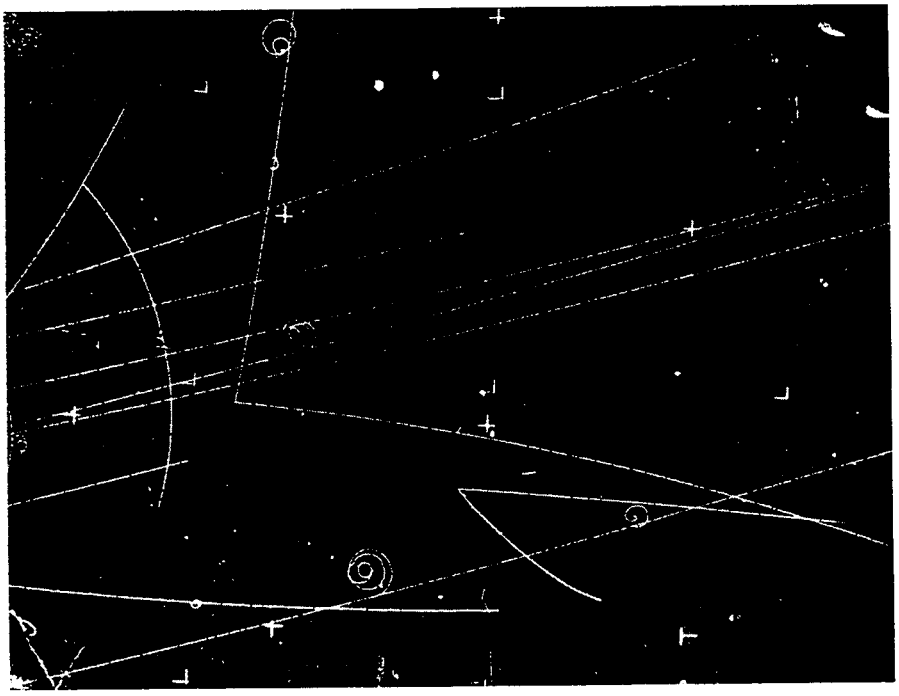


Fig.1. $\pi^- + p \rightarrow K^0 + \Lambda$.

emulsion, which had been so spectacularly successful in the hands of Powell's group, depended on the contiguous nature of the successive tracks at a production or decay vertex. The presence of neutral and therefore nonionizing particles between related charged particles, plus lack of even a rudimentary time resolution, made nuclear-emulsion techniques virtually unusable in this new field. The two known types of cloud chambers appeared to have equally insurmountable difficulties. The older Wilson expansion chamber had two difficulties that rendered it unsuitable for the job: if used at atmospheric pressure, its cycling period was measured in minutes, and if one increased its pressure to compensate for the long mean free path of nuclear interactions, its cycling period increased at least as fast as the pressure was increased. Therefore the number of observed reactions per day started at an almost impossibly low value, and dropped as «corrective action» was taken. The diffusion cloud chamber was plagued by «background problems», and had an additional disadvantage–its sensitive volume was confined in the vertical direction to a height of only a few centimeters. What we concluded from all this was simply that particle physicists needed a track-recording device with solid or liquid density (to increase the rate of production of nuclear events by a factor of 100), with uniform sensitivity (to avoid the problems of the sensitive layer in the diffusion chamber), and with fast cycling time (to avoid the Wilson chamber problems). And of course, any cycling detector would permit the association of charged tracks joined by neutral tracks, which was denied to the user of nuclear emulsion.

In late April of 1953 I paid my annual visit to Washington, to attend the meeting of the American Physical Society. At luch on the first day, I found myself seated at a large table in the garden of the Shoreham Hotel. All the seats but one were occupied by old friends from World War II days, and we reminisced about our experiences at the M.I.T. radar laboratory and at Los Alamos. A young chap who had not experienced those exciting days was seated at my left, and we were soon talking of our interests in physics. He expressed concern that no one would hear his 10-min contributed paper, because it was scheduled as the final paper of the Saturday afternoon session, and therefore the last talk to be presented at the meeting. In those days of slow airplanes, there were even fewer people in the audience for the last paper of the meeting than there are now–if that is possible. I admitted that I wouldn't be there, and asked him to tell me what he would be reporting. And that is how I heard first hand from Donald Glaser how he had invented the bubble chamber, and to what state he had brought its development. And of course he has since

described those achievements from this platform[39]. He showed me photographs of bubble tracks in a small glass bulb, about 1 centimeter in diameter and 2 centimeters long, filled with diethyl ether. He stressed the need for absolute cleanliness of the glass bulb, and said that he could maintain the ether in a superheated state for an average of many seconds before spontaneous boiling took place. I was greatly impressed by his work, and it immediately occurred to me that this could be the «big idea» I felt was needed in particle physics.

That night in my hotel room I discussed what I had learned with my colleague from Berkeley, Frank Crawford. I told Frank that I hoped we could get started on the development of a liquid hydrogen chamber, much larger than anything Don Glaser was thinking about, as soon as I returned to Berkeley. He volunteered to stop off in Michigan on the way back to Berkeley, which he did, and learned everything he could about Glaser's technique.

I returned to Berkeley on Sunday, May 1, and on the next day Lynn Stevenson started to keep a new notebook on bubble chambers. The other day, when he saw me writing this talk, he showed me that old notebook with its first entry dated May 2, 1953, with Van der Waals' equation on the first page, and the isotherms of hydrogen traced by hand onto the second page. Frank Crawford came home a few days later, and he and Lynn moved into the «student shop» in the synchrotron building, to build their first bubble chamber. They were fortunate in enlisting the help of John Wood, who was an accelerator technician at the synchrotron. The three of them put their first efforts into a duplication of Glaser's work with hydrocarbons. When they had demonstrated radiation sensitivity in ether, they built a glass chamber in a Dewar flask to try first with liquid nitrogen and then with liquid hydrogen.

I remember that on several occasions I telephoned to the late Earl Long at the University of Chicago, for advice on cryogenic problems. Dr. Long gave active support to the liquid hydrogen bubble chamber that was being built at that time by Roger Hildebrand and Darragh Nagle at the Fermi Institute in Chicago. In August of 1953 Hildebrand and Nagle[40] showed that superheated hydrogen boiled faster in the presence of a γ-ray source than it did when the source was removed. This is a necessary (though not sufficient) condition for successful operation of a liquid hydrogen bubble chamber, and the Chicago work was therefore an important step in the development of such chambers. The important unanswered question concerned the bubble density–was it sufficient to see tracks of «minimum ionizing» particles, or did liquid hydrogen (as my colleagues had just shown that liquid nitrogen did) produce bubbles but no visible tracks?

John Wood[41] saw the first tracks in a 1.5-inch-diameter liquid hydrogen bubble chamber in February of 1954. The Chicago group could certainly have done so earlier, by rebuilding their apparatus, but they switched their efforts to hydrocarbon chambers, and were rewarded by being the first physicists to publish experimental results obtained by bubble chamber techniques. Fig. 2 is a photograph of Wood's first tracks.

Fig. 2. First tracks in hydrogen.

At the Lawrence Radiation Laboratory, we have long had a tradition of close cooperation between physicists and technicians. The resulting atmosphere, which contributed so markedly to the rapid development of the liquid hydrogen bubble chamber, led to an unusual phenomenon: none of the scientific papers on the development of bubble-chamber techniques in my research group were signed by experimenters who were trained as physicists or who had had previous cryogenic experience. The papers all had authors who were listed on the Laboratory records as technicians, but of course the physicists concerned knew what was going on, and offered many suggestions. Nonethe-

less, our technical associates carried the main responsibility, and published their findings in the scientific literature. I believe this is a healthy change from practices that were common a generation ago; we all remember papers signed by a single physicist that ended with a paragraph saying, «I wish to thank Mr. ....., who built the apparatus and took much of the data».

And speaking of acknowledgments, John Wood's first publication, in addition to thanking Crawford, Stevenson, and me for our advice and help, said, «I am indebted to A. J. Schwemin for help with the electronic circuits.» «Pete» Schwemin, the most versatile technician I have ever known, became so excited by his initial contact with John Wood's 1.5-inch-diameter all-glass chamber that he immediately started the construction of the first metal bubble chamber with glass windows. All earlier chambers had been made completely of smooth glass, without joints, to prevent accidental boiling at sharp points; such boiling of course destroyed the superheat and made the chamber insensitive to radiation. Both Glaser and Hildebrand stressed the long times their liquids could be held in the superheated condition; Hildebrand and Nagle averaged 22 sec, and observed one superheat period of 70 sec. John Wood[41] reported, «We were discouraged by our inability to attain the long times of superheat, until the track photographs showed that it was not important in the successful operation of a large bubble chamber.» I have always felt that second to Glaser's discovery of tracks this was the key observation in the whole development of bubble-chamber technique. As long as one «expanded the chamber» rapidly, bubbles forming on the wall didn't destroy the superheated condition of the main volume of the liquid, and it remained sensitive as a track-recording medium.

Pete Schwemin, with the help of Douglas Parmentier[42], built the 2.5-inch-diameter hydrogen chamber in record time, as the world's first «dirty chamber». I've never liked that expression, but it was used for a while to distinguish chambers with windows gasketed to metal bodies from all-glass chambers. Because of its «dirtiness», the 2.5-inch chamber boiled at its walls, but still showed good tracks throughout its volume. Now that «clean» chambers are of historical interest only, we can be pleased that the modern chambers need no longer be stigmatized by the adjective «dirty».

Lynn Stevenson's notebook shows a diagram of John Wood's chamber dated January 25, 1954, with Polaroid pictures of tracks in hydrogen. A month later he recorded details of Schwemin's 2.5-inch chamber, and drew a complete diagram dated March 5. (That was the day after the *Physical Review* received Wood's letter announcing the first observation of tracks.)

On April 29, Schwemin and Parmentier photographed their first tracks; these are shown in Fig. 3. (Things were happening so fast at this time that the 2.5-inch system was never photographed as a whole before it ended up on the scrap pile.)

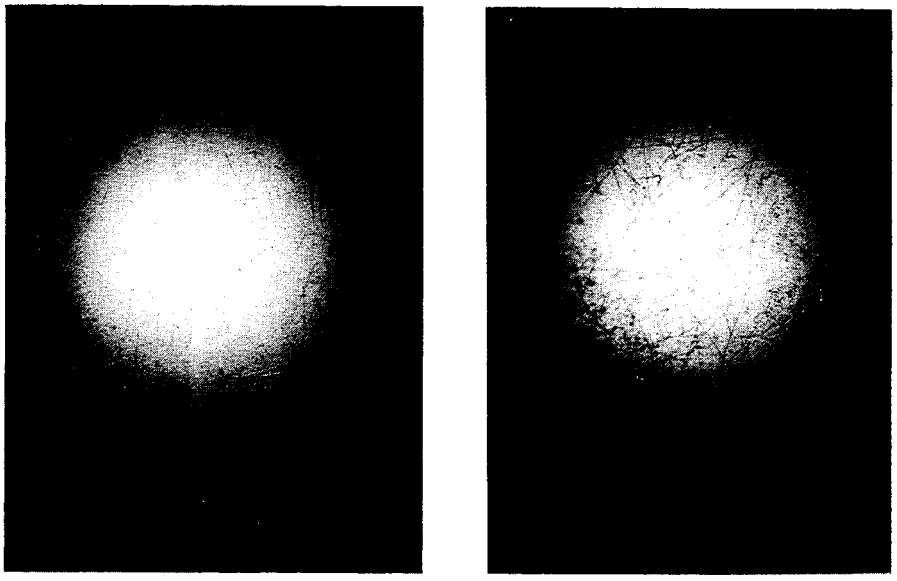


Fig. 3. Tracks in 2.5-inch chamber: neutrons (left); γ-rays (right).

In August, Schwemin and Parmentier separately built two different 4-inch-diameter chambers. Both were originally expanded by internal bellows, and Parmentier's 4-inch chamber gave tracks on October 6. Schwemin's chamber produced tracks three weeks later, and survived as *the* 4-inch chamber (see Fig. 4). The bellows systems in both chambers failed, but it turned out to be easier to convert Schwemin's chamber to the vapor expansion system that was used in all our subsequent chambers until 1962. (In that year, the 25-inch chamber introduced the «Ω-bellows» that is now standard for large chambers.)

Fig. 5 shows all our chambers displayed together a few weeks ago, at the request of Swedish Television. As you can see, we all look pretty pleased to see so many of our «old friends» side by side for the first time.

Fig. 6 shows an early picture of multiple meson production in the 4-inch chamber. This chamber was soon equipped with a pulsed magnetic field, and in that configuration it was the first bubble chamber of any kind to show

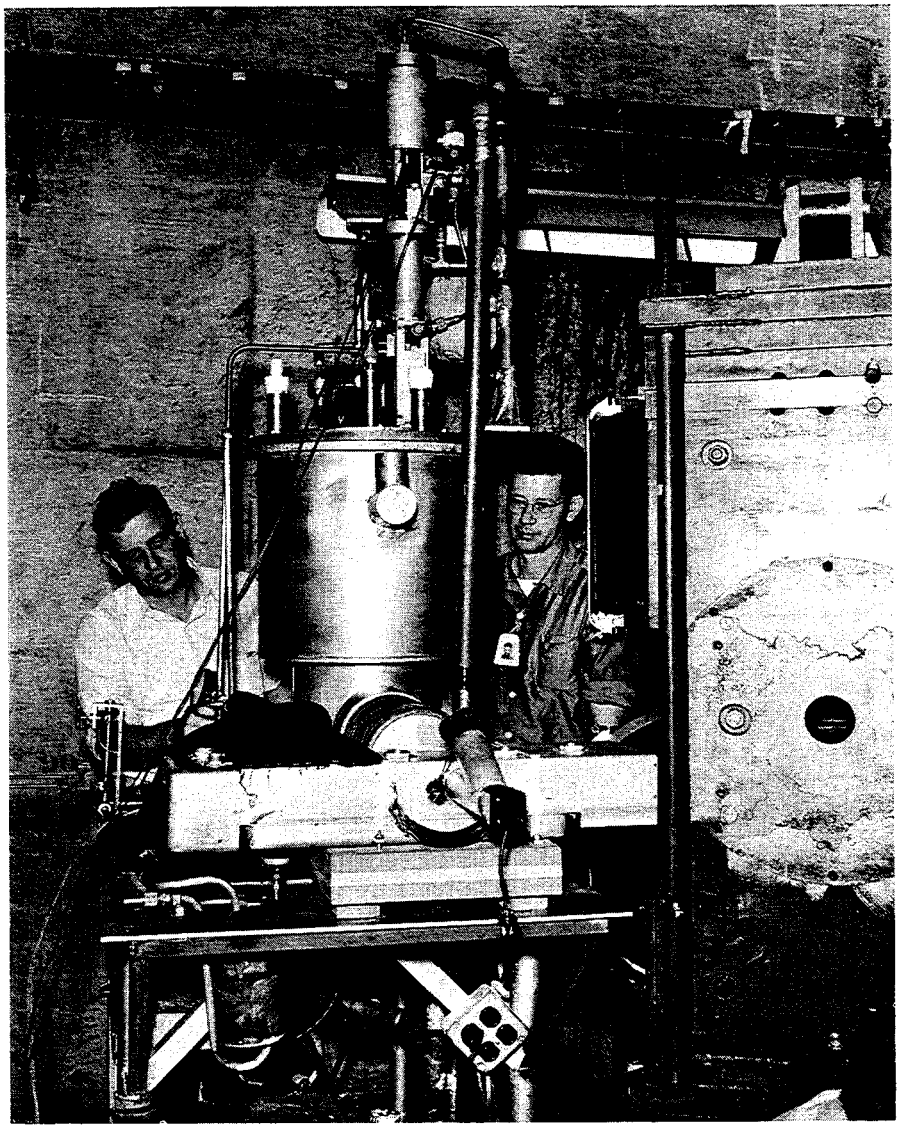


Fig. 4. The 4-inch chamber. D. Parmentier (left), A. J. Schwemin (right).

magnetically curved tracks. It was then set aside by our group as we pushed on to larger chambers. But it ended its career as a useful research tool at the Berkeley electron synchrotron, after almost two million photographs of 300-MeV Bremsstrahlung passing through it had been taken and analyzed by Bob Kenney *et al.*[43].

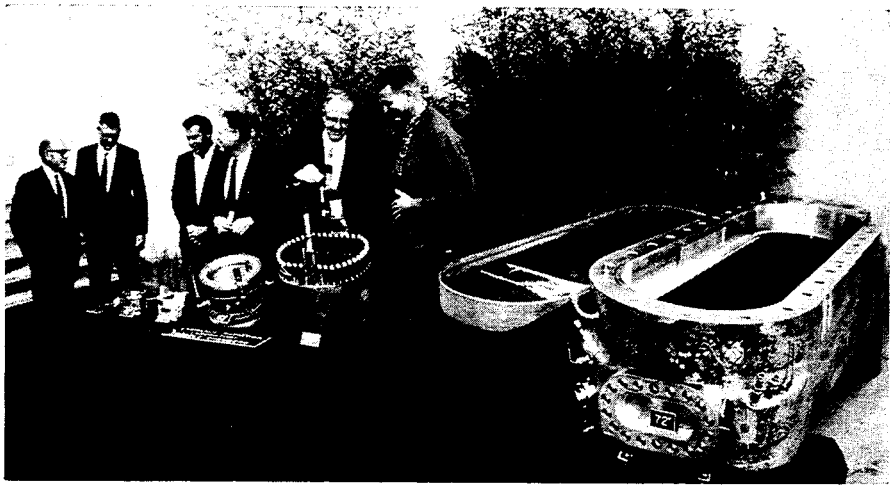


Fig. 5. Display of chambers, November 1968. From left to right, 1.5, 4, 6, 10, 15 and 72 inch chambers; Hernandez, Schwemin, Rinta, Watt, Alvarez and Eckman.

In the year 1954, as I have just recounted, various members of my research group had been responsible for the successful operation of four separate liquid hydrogen bubble chambers, increasing in diameter from 1.5 inches to 4 inches. By the end of that eventful year, it was clear that it would take a more concerted engineering-type approach to the problem if we were to progress to the larger chambers we felt were essential to the solution of high-energy physics problems. I therefore enlisted the assistance of three close associates, J. Donald Gow, Robert Watt and Richard Blumberg. Don Gow and Bob Watt had taken over full responsibility for the development and operation of the 32-MeV linear accelerator that had occupied all my attention from its inception late in 1945 until it first operated in late 1947. Neither of them had any experience with cryogenic techniques, but they learned rapidly, and were soon leaders in the new technology of hydrogen-bubble chambers. Dick Blumberg had been trained as a mechanical engineer, and he had designed the equipment used by Crawford, Stevenson and me in our experiments, then in progress, on the Compton scattering of γ-rays by protons[44].

Wilson Powell had built two large magnets to accommodate his Wilson Cloud Chambers, pictures from which adorned the walls of every cyclotron laboratory in the world. He very generously placed one of these magnets at our disposal, and Dick Blumberg immediately started the mechanical design of the 10-inch chamber – the largest size we felt could be accommodated in

the well of Powell's magnet. Blumberg's drafting table was in the middle of the single room that contained the desks of all the members of my research group. Not many engineers will tolerate such working conditions, but Blumberg was able to do so and he produced a design that was quickly built in the main machine shop. All earlier chambers had been built by the experimenters themselves. The design of the 10-inch chamber turned out to be a much larger job than we had foreseen. By the time it was completed, eleven members of the Laboratory's Mechanical Engineering Department had worked on it, including Rod Byrns and John Mark. The electrical engineering aspects of all our large chambers were formidable, and we are indebted to Jim Shand for his leadership in this work for many years.

Great difficulty was experienced with the first operation of the 10-inch

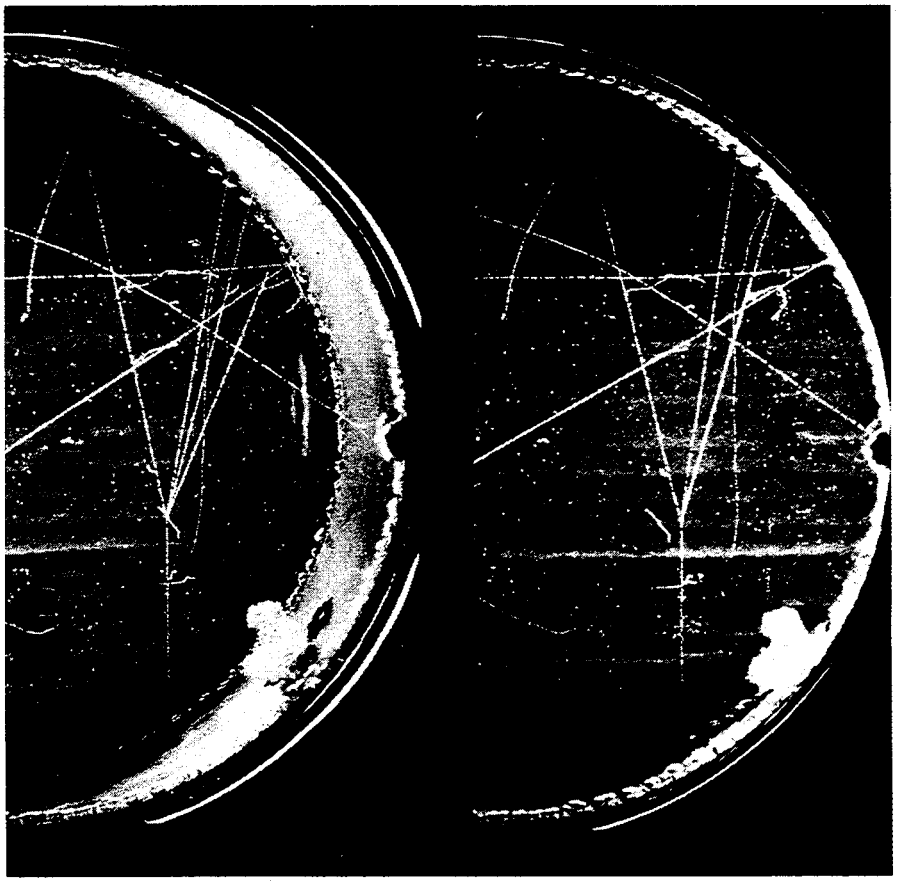


Fig. 6. Multiple meson production in 4-inch chamber.

chamber: too much hydrogen was vaporized at each «expansion». Pete Schwemin quickly diagnosed the trouble and built a fast-acting valve that permitted the chamber to be pulsed every 6 sec, to match the Bevatron's cycling time.

It would be appropriate to interrupt this description of the bubble-chamber development program to describe the important observations made possible by the operation of the 10-inch chamber early in 1956, but instead, I will preserve the continuity by describing the further development of the hardware. In December of 1954, shortly after the 4-inch chamber had been operated in the cyclotron building for the first time, it became evident to me that the 10-inch chamber we had just started to design wouldn't be nearly large enough to tell us what we wanted to know about the strange particles. The tracks of these objects had been photographed at Brookhaven[17], and we knew they were produced copiously by the Bevatron.

The size of the «big chamber» was set by several different criteria, and fortunately all of them could be satisfied by one design. (Too often, a designer of new equipment finds that one essential criterion can be met only if the objects is very large, while an equally important criterion demands that it be very small.) All «dirty chambers» so far built throughout the world had been cylindrical in shape, and were characterized by their diameter measurement. By studying the relativistic kinematics of strange particles produced by Bevatron beams, and more particularly by studying the decay of these particles, I convinced myself that the big chamber should be rectangular, with a length of at least 30 inches. This length was next increased to 50 inches in order that there would be adequate amounts of hydrogen upstream from the required decay region, in which production reactions could take place. Later the length was changed to 72 inches, when it was realized that the depth of the chamber could properly be less than its width and that the change could be made without altering the volume. The production region corresponded to about 10% of a typical pion–proton mean free path, and the size of the decay region was set by the relativistic time-dilated decay lengths of the strange particles, plus the requirement that there be a sufficient track length available in which to measure magnetic curvature in a «practical magnetic field» of 15000 gauss. In summary, then, the width and depth of the chamber came rather simply from an examination of the shape of the ellipses that characterize relativistic transformations at Bevatron energies, plus the fact that the magnetic field spreads the particles across the width but not along the depth of the chamber.

The result of this straightforward analysis was a rather frightening set of numbers: The chamber length was 72 inches; its width was 20 inches, and its depth was 15 inches. It had to be pervaded by a magnetic field of 15 000 gauss, so its magnet would weigh at least 100 tons and would require 2 or 3 megawatts to energize it. It would require a window 75 inches long by 23 inches wide and 5 inches thick to withstand the (deuterium) operating pressure of 8 atmospheres, exerting force of 100 tons on the glass. No one had any experience with such large volumes of liquid hydrogen; the hydrogen–oxygen rocket engines that now power the upper stages of the Saturn boosters were still gleams in the eyes of their designers–these were pre-Sputnik days. The safety aspects of the big chamber were particularly worrisome. Low temperature laboratories had a reputation for being dangerous places in which to work, and they didn't deal with such large quantities of liquid hydrogen, and what supplies they did use were kept at atmospheric pressure.

For some time, the glass-window problem seemed insurmountable – no one had ever cast and polished such a large piece of optical glass. Fortunately for the eventual success of the project, I was able to persuade myself that the chamber body could be constructed of a transparent plastic cylinder with metallic end plates. This notion was later demolished by my engineering colleagues, but it played an important role in keeping the project alive in my own mind until I was convinced that the glass window could be built. As an indication of the cryogenic «state of the art» at the time we worried about the big window, I can recall the following anecdote. One day, while looking through a list of titles of talks at a recent cryogenic conference, I spotted one that read, «Large glass window for viewing liquid hydrogen.» Eagerly I turned to the paper–but it described a metallic Dewar vessel equipped with a glass window 1 inch in diameter!

Don Gow was now devoting all his time to hydrogen bubble chambers, and in January of 1955 we interested Paul Hernandez in taking a good hard engineering look at the problems involved in building and housing the 72-inch bubble chamber. We were also extremely fortunate in being able to interest the cryogenic engineers at the Boulder, Colorado, branch of the National Bureau of Standards in the project. Dudley Chelton, Bascomb Birmingham and Doug Mann spent a great deal of time with us, first educating us in large-scale liquid-hydrogen techniques, and later cooperating with us in the design and initial operation of the big chamber.

In April of 1955, after several months of discussion of the large chamber, I wrote a document entitled «The Bubble Chamber Program at UCRL».

This paper showed in some detail why it was important to build the large chamber, and outlined a whole new way of doing high-energy physics with such a device. It stressed the need for semiautomatic measuring devices (which had not previously been proposed), and described how electronic computers would reconstruct tracks in space, compute momenta, and solve problems in relativistic mechanics. All these techniques are now part of the «standard bubble-chamber method», but in April of 1955 no one had yet applied them. Of all the papers I have written in my life, none gives me so much satisfaction on rereading as does this unpublished prospectus.

After Paul Hernandez and Don Gow had estimated that the big chamber, including its building and power supplies, would cost about 2.5 million dollars, it was clear that a special AEC appropriation was required; we could no longer build our chambers out of ordinary laboratory operating money. In fact, the document I've just described was written as a sort of proposal to the AEC for financial support–but without mentioning money! I asked Ernest Lawrence if he would help me in requesting extra funds from the AEC. He read the document, and agreed with the points I had made. He then asked me to remind him of the size of the world's largest hydrogen chamber. When I replied that it was 4 inches in diameter, he said he thought I was making too large an extrapolation in one step, to 72 inches. I told him that the 10-inch chamber was on the drawing board, and if we could make it work, the operation of the 72-inch chamber was assured. (And if we couldn't make it work, we could refund most of the 2.5 million.) This wasn't obvious until I explained the hydraulic aspects of the expansion system of the 72-inch chamber; it was arranged so that the 20-inch wide, 72-inch long chamber could be considered to be a large collection of essentially independently expanded 10-inch square chambers. He wasn't convinced of the wisdom of the program, but in a characteristic gesture, he said, «I don't believe in your big chamber, but I do believe in you, and I'll help you to obtain the money.» I therefore accompanied him on his next trip to Washington, and we talked in one day to three of the five Commissioners: Lewis Strauss, Willard Libby (who later spoke from this podium), and the late John Von Neumann, the greatest mathematical physicist then living. That evening, at a cocktail party at Johnny Von Neumann's home, I was told that the Commission had voted that afternoon to give the laboratory the 2.5 million dollars we had requested. All we had to do now was build the thing and make it work!

Design work had of course been under way for some time, but it was now rapidly accelerated. Don Gow assumed a new role that is not common in

ing period of more than a year and a half before another τ meson showed up.

In 1951, the year after the τ meson and the V particles were finally seen again, O'Ceallaigh[25] observed the first of his kappa mesons in nuclear emulsion. Each such event involved the decay at rest of a heavy meson into a muon with a different energy. We now know these particles as K^+ mesons decaying into $\mu^+ + \pi^0 + \nu$, so the explanation of the broad muon energy spectrum is now obvious. But it took some time to understand this in the early 1950's, when these particles appeared one by one in different laboratories. In 1953, Menon and O'Ceallaigh[26] found the first $K_{\pi 2}$ or θ meson, with a decay into $\pi^+ + \pi^0$. The identification of the θ and τ mesons as different decay modes of the same K meson is one of the great stories of particle physics, and it will be mentioned later in this lecture.

The identification of the neutral Λ emerged from the combined efforts of the cosmic-ray cloud-chambers groups, so I won't attempt to assign credit for its discovery. But it does seem clear that Thompson *et al.*[27] were the first to establish the decay scheme of what we now know as the K_1^0 meson: $K_1^0 \rightarrow \pi^+ + \pi^-$. The first example of a charged Σ hyperon was seen in emulsion by the Genoa and Milan groups[28], in 1953. And after that, the study of strange particles passed, to a large extent, from the cosmic-ray groups to the accelerator laboratories.

So by the time the Bevatron first operated, in 1954, a number of different strange particles had been identified: several charged particles and a neutral one all with masses in the neighborhood of 500 MeV, and three kinds of particles heavier than the proton. In order of increasing mass, these were the neutral Λ, the two charged Σ's (plus and minus), and the negative cascade (Ξ^-), which decayed into a Λ and a negative pion.

The strange particles all had lifetimes shorter than any known particles except the neutral pion. The hyperons all had lifetimes of approximately 10^{-10} sec, or less than 1% of the charged pion lifetime. When I say that they were called strange particles because their observed lifetimes presented such a puzzle for theoretical physicists to explain, I can imagine the lay members in this audience saying to themselves, « Yes, I can't see how anything could come apart so fast. » But the strangeness of the strange particles is not that they decay so rapidly, but that they last almost a million million times longer than they should–physicists couldn't explain why they didn't come apart in about 10^{-21} sec.

I won't go into the details of the dilemma, but we can note that a similar problem faced the physics community when the muon was found to be so

physics laboratories, but is well known in military organizations; he became my «chief of staff». In this position, he coordinated the efforts of the physicists and engineers; he had full responsibility for the careful spending of our precious 2.5 million dollars, and he undertook to become an expert second to none in all the technical phases of the operation, from low-temperature thermodynamics to safety engineering. His success in this difficult task can be recognized most easily in the success of the whole program, culminating in the fact that I am speaking here this afternoon. I am sorry that Don Gow can't be here today; he died several years ago, but I am reminded of him every day – my three-year-old son is named Donald in his memory.

The engineering team under Paul Hernandez's direction proceeded rapidly with the design, and in the process solved a number of difficult problems in ways that have become standard «in the industry». A typical problem involved the very considerable differential expansion between the stainless steel chamber and the glass window. This could be lived with in the 10-inch chamber, but not in the 72-inch. Jack Franck's «inflatable gasket» allowed the glass to be seated against the chamber body only after both had been cooled to liquid hydrogen temperature.

Just before leaving for Stockholm, I attended a ceremony at which Paul Hernandez was presented with a trophy honoring him as a «Master Designer» for his achievements in the engineering of the 72-inch chamber. I had the pleasure of telling in more detail than I can today of his many contributions to the success of our program. One of his associates recalled a special service that he rendered not only to our group but to all those who followed us in building liquid hydrogen-bubble chambers. Hernandez and his associates wrote a series of «Engineering Notes», on matters of interest to designers of hydrogen-bubble chambers, that soon filled a series of notebooks that spanned 3 feet of shelf space. Copies of these were sent to all interested parties on both sides of the Atlantic, and I am sure that they resulted in a cumulative savings to all bubble-chamber builders of several million dollars; had not all this information been readily available, the test programs and calculations of our engineering group would have required duplication at many laboratories, at a large expense of money and time. Our program moved so rapidly that there was never time to put the Engineering Notes into finished form for publication in the regular literature. For this reason, one can now read review articles on bubble-chamber technology, and be quite unaware of the part that our Laboratory played in its development. There are no references to papers by members of our group, since those papers were never written – the data that

would have been in them had been made available to everyone who needed them at a much earlier date.

And just to show that I was also deeply involved in the chamber design, I might recount how I purposely «designed myself into a corner» because I thought the results were important, and I thought I could invent a way out of a severe difficulty, if given the time. All previous chambers had had two windows, with «straight through» illumination. Such a configuration reduces the attainable magnetic field, because the existence of a rear pole piece would interfere with the light-projection system. I made the decision that the 72-inch chamber would have only a top window, thereby permitting the magnetic field to be increased by a lower pole piece and at the same time saving the cost of the extra glass window, and also providing added safety by eliminating the possibility that liquid hydrogen could spill through a broken lower window. The only difficulty was that for more than a year, as the design was firmed up and the parts were fabricated, none of us could invent a way both to illuminate and to photograph the bubbles through the same window. Duane Norgren, who has been responsible for the design of all our bubble-chamber cameras, discussed the matter with me at least once a week in that critical year, and we tried dozens of schemes that didn't quite do the job. But as a result of our many failures, we finally came to understand all the problems, and we eventually hit on the retrodirecting system known as coat hangers. This solution came none too soon; if it had been delayed by a month or more, the initial operation of the 72-inch chamber would have been correspondingly delayed. We took many other calculated risks in designing the system; if we had postponed the fabrication of the major hardware until we had solved all the problems on paper, the project might still not be completed. Engineers are conservative people by nature; it is the ultimate disgrace to have a boiler explode or a bridge collapse. We were therefore fortunate to have Paul Hernandez as our chief engineer; he would seriously consider anything his physics colleagues might suggest, no matter how outlandish it might seem at first sight. He would firmly reject it if it couldn't be made safe, but before rejecting any idea for lack of safety he would use all the ingenuity he possessed to make it safe.

We felt that we needed to built a test chamber to gain experience with a single-window system, and to learn to operate with a hydrogen refrigerator; our earlier chambers had all used liquid hydrogen as a coolant. We therefore built and operated the 15-inch chamber in the Powell magnet, in place of the 10-inch chamber that had served us so well.

The 72-inch chamber operated for the first time on March 24, 1959, very nearly four years from the time it was first seriously proposed. Fig. 7 shows it at about that time. The «start-up team» consisted of Don Gow, Paul Hernandez and Bob Watt, all of whom had played key roles in the initial operation

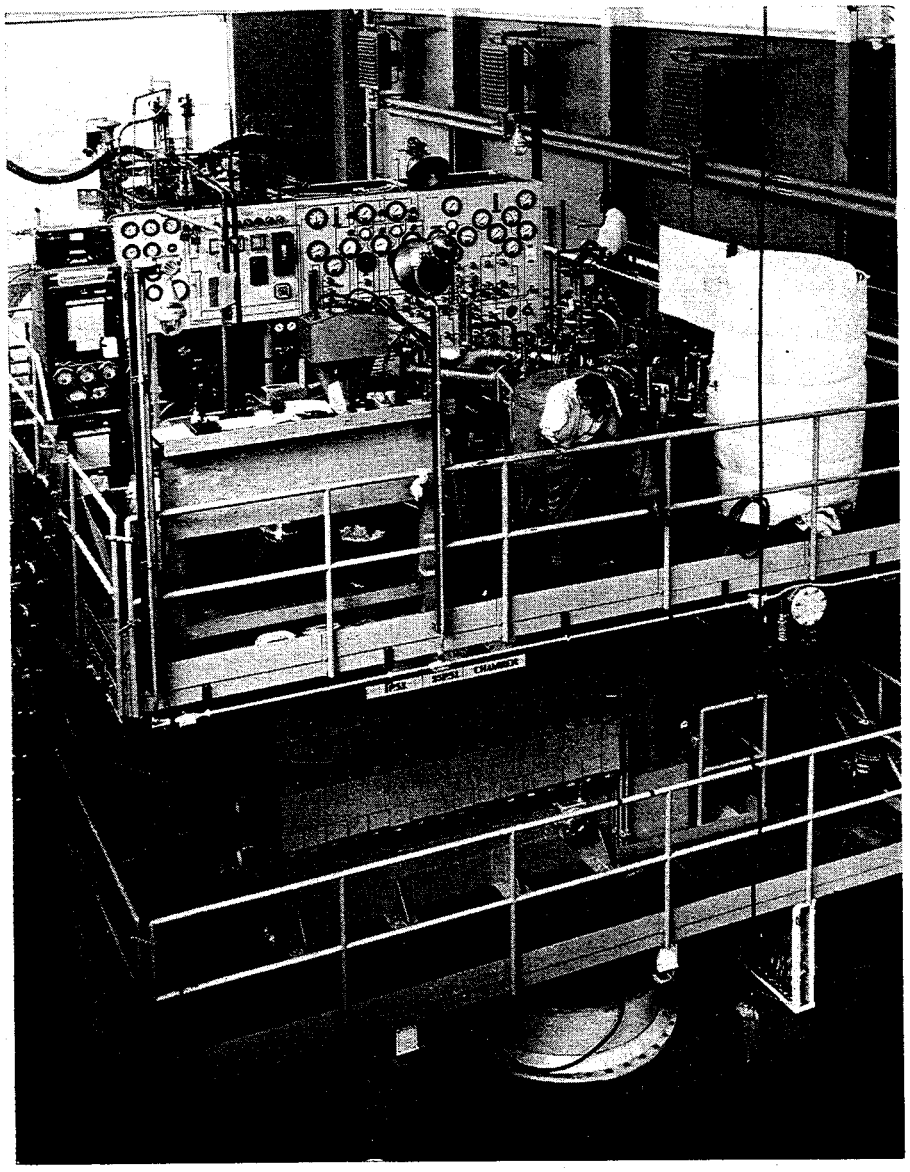


Fig. 7. The 72-inch bubble chamber in its building.

of the 15-inch chamber. Bob Watt and Glenn Eckman have been responsible for the operation of all our chambers from the earliest days of the 10-inch chamber, and the success of the whole program has most often rested in their hands. They have maintained an absolutely safe operating record in the face of very severe hazards, and they have supplied their colleagues in the physics community with approximately ten million high-quality stereo photographs. And most recently, they have shown that they can design chambers as well as they have operated them. The 72-inch chamber was recently enlarged to an 82-inch size, incorporating to a large extent the design concepts of Watt and Eckman.

Although I haven't done justice to the contributions of many close friends and associates who shared in our bubble-chamber development program, I must now turn to another important phase of our activities – the data-analysis program. Soon after my 1955 prospectus was finished, Hugh Bradner undertook to implement the semiautomatic measuring machine proposal. He first made an exhaustive study of commercially available measuring machines, encoding techniques, etc., and then, with Jack Franck, designed the first «Franckenstein». This rather revolutionary device had been widely copied, to such an extent that objects of its kind are now called «conventional» measuring

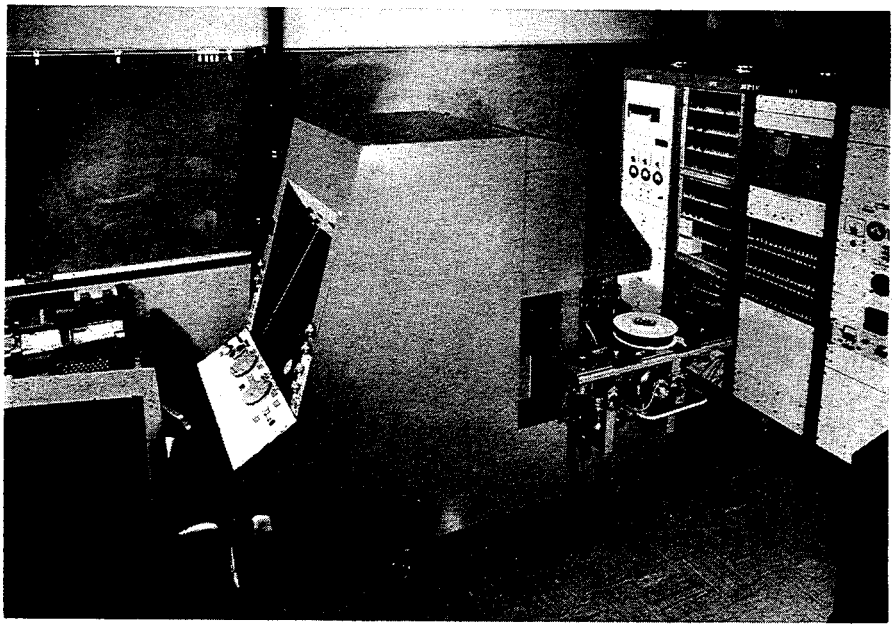


Fig. 8. «Franckenstein».

machines (Fig. 8). Our first Franckenstein was operating reliably in 1957, and in the summer of 1958 a duplicate was installed in the U. S. exhibit at the «Atoms for Peace» exposition in Geneva. It excited a great deal of interest in the high-energy physics community, and a number of groups set out to make similar machines based on its design. Almost everyone thought at first that our provision for automatic track following was a needless waste of money, but over the years, that feature has also come to be «conventional».

Jack Franck then went on to design the Mark II Franckenstein, to measure 72-inch bubble-chamber film. He had the first one ready to operate just in time to match the rapid turn-on of the big chamber, and he eventually built three more of the Mark II's. Other members of our group then designed and perfected the faster and less expensive S M P system, which added significantly to our «measuring power». The moving forces in this development were Pete Schwemin, Bob Hulsizer, Peter Davey, Ron Ross and Bill Humphrey[45]. Our final and most rewarding effort to improve our measuring ability was fulfilled several years ago, when our first Spiral Reader became operational. This single machine has now measured more than one and a half million high-energy interactions, and has, together with its almost identical twin, measured one and a quarter million events in the last year. The SAAB Company here in Sweden is now building and selling Spiral Readers to European laboratories.

The Spiral Reader had a rather checkered career, and it was on several occasions believed by most workers in the field to have been abandoned by our group. The basic concept of the spiral scan was supplied by Bruce McCormick, in 1956. Our attempts to reduce his ideas to practice resulted in failure, and shortly after that, McCormick moved to Illinois, where he has since been engaged in computer development. As the cost of transistorized circuits dropped rapidly in the next years, we tried a second time to implement the Spiral Reader concept, using digital techniques to replace the analog devices of the earlier machine. The second device showed promise, but its «hard-wired logic» made it too inflexible, and the unreliability of its electronic components kept it undergoing repair most of the time. The mechanical and optical components of the second Spiral Reader were excellent, and we hated to drop the whole project simply because the circuitry didn't come up to the same standard. In 1963, Jack Lloyd suggested that we use one of the new breed of small high-speed, inexpensive computers to supply the logic and the control circuits for the Spiral Reader. He then demonstrated great qualities of leadership by delivering to our research group a machine that has performed even better than he had promised it would. In addition to his development of the hard-

ware, he initiated POOH, the Spiral Reader filtering program, which was brought to a high degree of perfection by Jim Burkhard. The smooth and rapid transition of the Spiral Reader from a developmental stage into a useful operational tool was largely the result of several years of hard work on the part of Gerry Lynch and Frank Solmitz. Fig. 9, from a talk I gave two and a half years ago[46], shows how the measuring power of our group has increased over the years, with only a modest increase in personnel.

According to a simple extrapolation of the exponential curve we had been

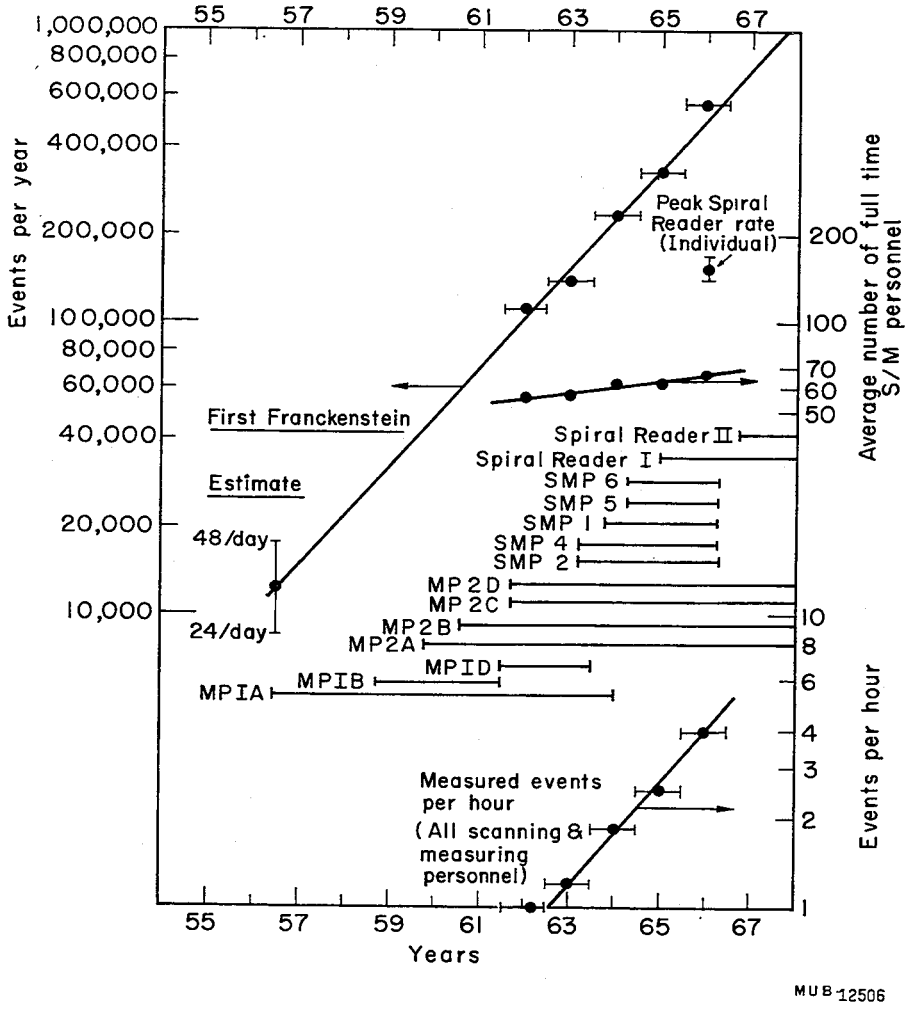


Fig. 9. Measuring rates.

on from 1957 through 1966, we would expect to be measuring 1.5 million events per year some time in 1969. But we have already reached that rate and we will soon be leveling off about there because we have stopped our development work in this area.

The third key ingredient of our development program has been the continually increasing sophistication in our utilization of computers, as they have increased in computational speed and memory capacity. While I can speak from a direct involvement in the development of bubble chambers and measuring machines, and in the physics done with those tools, my relationship to our computer programming efforts is largely that of an amazed spectator. We were most fortunate that in 1956 Frank Solmitz elected to join our group. Although the rest of the group thought of themselves as experimental physicists, Solmitz had been trained as a theorist, and had shown great aptitude in the development of statistical methods of evaluating experimental data. When he saw that our first Franckenstein was about to operate, and no computer programs were ready to handle the data it would generate, he immediately set out to remedy the situation. He wrote HYDRO, our first system program for use on the IBM 650 computer. In the succeeding twelve years he has continued to carry the heavy responsibility for all our programming efforts. A major breakthrough in the analysis of bubble-chamber events was made in the years 1957 through 1959. In this period, Solmitz and Art Rosenfeld, together with Horace Taft from Yale University and Jim Snyder from Illinois, wrote the first «fitting routine», GUTS, which was the core of our first «kinematics program», KICK. To explain what KICK did, it is easiest to describe what physicists had to do before it was written. HYDRO and its successor, PANG, listed for each vertex the momentum and space angles of the tracks entering or leaving that vertex, together with the calculated errors in these measurements. A physicist would plot the angular coordinates on a stereographic projection of a unit sphere known as a Wolff-plot. If he was dealing with a three-track vertex–and that was all we could handle in those days–he would move the points on the sphere, within their errors, if possible, to make them coplanar. And of course he would simultaneously change the momentum values, within their errors, to insure that the momentum vector triangle closed, and energy was conserved. Since momentum is a vector quantity, the various conditions could be simultaneously satisfied only after the angles and the absolute values of the momenta had been changed a number of times in an iterative procedure. The end result was a more reliable set of momenta and angles, constrained to fit the conservation laws of energy and

momentum. In a typical case, an experienced physicist could solve only a few Wolff-plot problems in a day. (Lynn Stevenson had written a specific program, COPLAN, that solved a particular problem of interest to him that was later handled by the more versatile GUTS.)

GUTS was being written at a time when one highly respected visitor to the group saw the large pile of PANG printout that had gone unanalyzed because so many of our group members were writing GUTS–a program that was planned to do the job automatically. Our visitor was very upset at what he told me was a «foolish deployment of our forces». He said, «If you would only get all those people away from their program writing, and put them to work on Wolff-plots, we'd have the answer to some really important physics in a month or two». I said I was sure we'd end up with at lot more physics in the next years if my colleagues continued to write GUTS and KICK. I'm sure that those who wrote these pioneering «fitting and kinematics programs» were subjected to similar pressures. Everyone in the high-energy physics community has long been indebted to these farsighted men because they knew that what they were doing was right. KICK was soon developed so that it gave an overall fit to several interconnected vertices, with various hypothetical identities of the several tracks assumed in a series of attempts at a fit. The relationship between energy and momentum depends on mass, so a highly constrained fit can be obtained only if the particle responsible for each track is properly identified. If the degree of constraint is not so high, more than one «hypothesis» (set of track identifications) may give a fit, and the physicist must use his judgment in making the identification.

As another example in this all-too-brief sketch of the computational aspects of our work, I will mention an important program, initiated by Art Rosenfeld and Ron Ross, that has removed much of the remaining drudgery from the bubble-chamber physicists' life. SUMX is a program that can easily be instructed to search quickly through large volumes of «kinematics program output», printing out summaries and tabulations of interesting data. (Like all our pioneering programs, SUMX was replaced by an improved and more versatile program – in this case, KIOWA. But I will continue to talk as though SUMX were still used.) A typical SUMX printout will be a computer-printed document 3 inches thick, with hundreds of histograms, scatter plots, etc.

Hundreds of histograms are similarly printed showing numbers of events with effective masses for many different combinations of particles, with various «cuts» on momentum transfer, etc. What all this amounts to is simply

that a physicist is no longer rewarded for his ability in deciding what histograms he should tediously plot and then examine. He simply tells the computer to plot all histograms of any possible significance, and then flips the pages to see which ones have interesting features.

One of my few real interactions with our programming effort came when I suggested to Gerry Lynch the need for a program he wrote that is known as GAME. In my work as a nuclear physicist before World War II, I had often been skeptical of the significance of the «bumps» in histograms, to which importance was attached by their authors. I developed my own criteria for judging statistical significance, by plotting simulated histograms, assuming curves to be smooth; I drew several samples of «Monte Carlo distributions», using a table of random numbers as the generator of the samples. I usually found that my skepticism was well founded because the «faked» histograms showed as much structure as the published ones. There are of course many statistical tests designed to help one evaluate the reality of bumps in histograms, but in my experience nothing is more convincing than an examination of a set of simulated histograms from an assumed smooth distribution.

GAME made it possible, with the aid of a few control cards, to generate a hundred histograms similar to those produced in any particular experiment. All would contain the same number of events as the real experiment, and would be based on a smooth curve through the experimental data. The standard procedure is to ask a group of physicists to leaf through the 100 histograms – with the experimental histogram somewhere in the pile – and vote on the apparent significance of the statistical fluctuations that appear. The first time this was tried, the experimenter – who had felt confident that his bump was significant – didn't know that his own histogram was in the pile, and didn't pick it out as convincing; he picked out two of the computer-generated histograms as looking significant, and pronounced all others – including his own – as of no significance! In view of this example, one can appreciate how many retractions of discovery claims have been avoided in our group by the liberal use of the GAME program.

As a final example from our program library, I'll mention FAKE, which, like SUMX, has been widely used by bubble-chamber groups all over the world. FAKE, written by Gerry Lynch, generates simulated measurements of bubble-chamber events to provide a method of testing the analysis programs to determine how frequently they arrive at an incorrect answer.

Now that I have brought you up to date on our parallel developments of hardware and software (computer programs), I can tell you what rewards we

have reaped, as physicists, from their use. The work we did with the 4-inch chamber at the 184-inch cyclotron and at the Bevatron cannot be dignified by the designation «experiment», but it did show examples of π–μ–e decay and neutral strange-particle decay. The experiences we had in scanning the 4-inch film merely whetted our appetite for the exciting physics we felt sure would be manifest in the 10-inch chamber, when it came into operation in Wilson Powell's big magnet.

Robert Tripp joined the group in 1955, and as his first contribution to our program he designed a «separated beam» of negative K mesons that would stop in the 10-inch chamber. We had two different reasons for starting our bubble-chamber physics program with observations of the behavior of K^- mesons stopping in hydrogen. The first reason involved physics: The behavior of stopping π^- mesons in hydrogen had been shown by Panofsky[47] and his co-workers to be a most fruitful source of fundamental knowledge concerning particle physics. The second reason was of an engineering nature: Only one Bevatron «straight section» was available for use by physicists, and it was in constant use. In order not to interfere with other users, we decided to set the 10-inch chamber close to a curved section of the Bevatron, and use secondary particles, from an internal target, that penetrated the wall of the vacuum chamber and passed between neighboring iron blocks in the return yoke of the Bevatron magnet. This physical arrangement gave us negative particles (K^- and π^- mesons) of a well-defined low momentum. By introducing an absorber into the beam, we brought the K^- mesons almost to rest, but allowed the lighter π^- mesons to retain a major fraction of their original momentum. The Powell magnet provided a second bending that brought the K^- mesons into the chamber, but kept the π^- mesons out. That was the theory of this first separated beam for bubble-chamber use. But in practice, the chamber was filled with tracks of pions and muons, and we ended up with only one stopped K^- per roll of 400 stereo pairs. It is now common for experimenters to stop one million K^- mesons in hydrogen, in a single experimental run, but the 137 K^- mesons we stopped in 1956[48] gave us a remarkable preview of what has now been learned in the much longer exposures. We measured the relative branching of $K^- + p$ into

$$\Sigma^- + \pi^+ : \Sigma^+ + \pi^- : \Sigma^0 + \pi^0 : \Lambda + \pi^0$$

And in the process, we made a good measurement of the Σ^0 mass. We plotted the first decay curves for the Σ^+ and Σ^- hyperons, and we observed for the first time the interactions of Σ^- hyperons and protons at rest. We felt amply

rewarded for our years of developmental work on bubble chambers by the very interesting observations we were now privileged to make.

We had a most exciting experience at this time, that was the result of two circumstances that no longer obtain in bubble-chamber physics. In the first place, we did all our own scanning of the photographic film. Such tasks are now carried out by professional scanners, who are carefully trained to recognize and record «interesting events». We had no professional scanners at the time, because we wouldn't have known how to train them before this first film became available. And even if they had been trained, we would not have let them look at the film–we found it so completely absorbing that there was always someone standing behind a person using one of our few film viewers, ready to take over when the first person's eyes tired. The second circumstance that made possible the accidental discovery I am about to describe was the very poor quality of our separated K^- beam–by modern standards. Most of the tracks we observed were made by negative pions or muons, but we also saw many positively charged particles–protons, pions and muons.

At first we kept no records of any events except those involving strange particles; we would look quickly at each frame in turn, and shift to the next one if no «interesting event» showed up. In doing this scanning, we saw many examples of π^+–μ^+–e^+ decays, usually from a pion at rest, and we soon learned about how long to expect the μ^+ track to be–about 1 centimeter. I did my scanning on a stereo viewer, so I probably had a better feeling for the length of a μ^+ track in space than did my colleagues, who looked at two projections of the stereo views, sequentially. Don Gow, Hugh Bradner, and I often scanned at the same time, and we showed each other whatever interesting events came into view. Each of us showed the others examples of what we thought was an unusual decay scheme: $\pi^- \rightarrow \mu^- \rightarrow e^-$. The decay of a μ^- at rest into an e^-, in hydrogen, was expected from the early observations by Conversi *et al.*[3], but Panofsky[47] had shown that a π^- meson couldn't decay at rest in hydrogen. Our first explanation for our observations was simply that the pion had decayed just before stopping. But we gradually became convinced that this explanation really didn't fit the facts. There were too many muon tracks of about the same length, and none that were appreciably longer or shorter, as the decay-in-flight hypothesis would predict. We now began to keep records of these «anomalous decays», as we still called them, and we found occasional examples in which the muon was horizontal in the chamber, so its length could be measured. (We had as yet no way of reconstructing tracks in space from two stereo views.) By comparing the measured length of the neg-

ative muon track with that of its more normal positive counterpart, we estimated that the negative muons had an energy of 5.4 MeV, rather than the well-known positive muon energy (from positive pion decay at rest) of 4.1 MeV. This confirmed our earlier suspicion that the long primary negative track couldn't be that of a pion, but it left us just as much in the dark as to the nature of the primary.

After these observations had been made, I gave a seminar describing what we had observed, and suggesting that the primary might be a previously unknown weakly interacting particle, heavier than the pion, that decayed into a muon and a neutral particle, either neutrino or photon. We had just made the surprising observation, shown in Fig. 10, that there was often a gap, measured in millimeters, between the end of the primary and the beginning of the secondary. This finding suggested diffusion by a rather long-lived negative particle that orbited around and neutralized one of the protons in the liquid hydrogen. We had missed many tracks with these «gaps» because no one had seen such a thing before; we simply ignored such track configurations by subconsciously assuming that they were unassociated events in a badly cluttered bubble chamber.

One evening, one of the members of our research team, Harold Ticho from our Los Angeles campus, was dining with Jack Crawford, a Berkeley astrophysicist he had known when they were students together. They discussed our observations at some length, and Crawford suggested the possibility that a fusion reaction might somehow be responsible for the phenomenon. They calculated the energy released in several such reactions, and found that it agreed with experiment if a stopped muon were to be binding together a proton and a deuteron into an HD μ^--molecular ion. In such a «molecule» the proton and deuteron would be brought into such close proximity for such a long time that they would fuse into ^{3}He, and could deliver their fusion energy to the muon by the process of internal conversion. However, Ticho and Crawford couldn't think of any mechanism that would make the reaction happen so often – the fraction of deuterons in liquid hydrogen is only 1 in 5000. They had, however, correctly identified the reaction, but a key ingredient in the theoretical explanation was still missing.

The next day, when we had all accepted the idea that stopped muons were catalyzing the fusion of protons and deuterons, our whole group paid a visit to Edward Teller, at his home. After a short period of introduction to the observations and to the proposed fusion reaction, he explained the high probability of the reaction as follows: the stopped muon radiated its way into

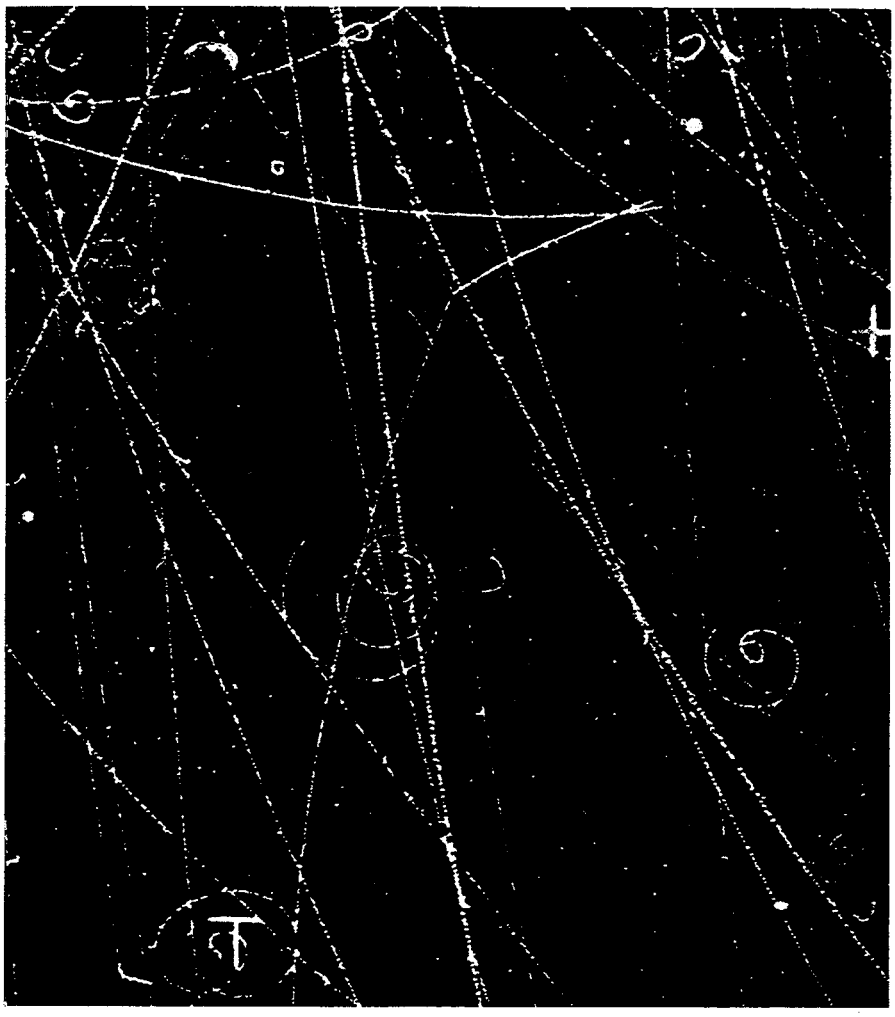


Fig. 10. Muon catalysis (with gap).

the lowest Bohr orbit around a proton. The resulting muonic hydrogen atom, $p\mu^-$, then had many of the properties of a neutron, and could diffuse freely through the liquid hydrogen. When it came close to the deuteron in an HD molecule, the muon would transfer to the deuteron, because the ground state of the $\mu^- d$ atom is lower than that of the $\mu^- p$ atom, in consequence of «reduced mass» effect. The new «heavy neutron» $d\mu^-$ might then recoil some distance as a result of the exchange reaction, thus explaining the «gap». The final stage

of capture of a proton into a $pd\mu^-$ molecular ion was also energetically favorable, so a proton and deuteron could now be confined close enough together by the heavy negative muon to fuse into a ^{3}He nucleus plus the energy given to the internally converted muon.

We had a short but exhilarating experience when we thought we had solved all of the fuel problems of mankind for the rest of time. A few hasty calculations indicated that in liquid HD a single negative muon would catalyze enough fusion reactions before it decayed to supply the energy to operate an accelerator to produce more muons, with energy left over after making the liquid HD from sea water. While everyone else had been trying to solve this problem by heating hydrogen plasmas to millions of degrees, we had apparently stumbled on the solution, involving very low temperatures instead. But soon, more realistic estimates showed that we were off the mark by several orders of magnitude–a «near miss» in this kind of physics!

Just before we published our results[49], we learned that the «μ-catalysis» reaction had been proposed in 1947 by Frank[50] as an alternative explanation of what Powell *et al.* had assumed (correctly) to be the decay of π^+ to μ^+. Frank suggested that it might be the reaction we had just seen in liquid hydrogen, starting with a μ^-, rather than with a π^+. Zel'dovitch[51] had extended the ideas of Frank concerning this reaction, but because their papers were not known to anyone in Berkeley, we had a great deal of personal pleasure that we otherwise would have missed.

I will conclude this episode by noting that we immediately increased the deuterium concentration in our liquid hydrogen and observed the expected increase in fusion reaction, and saw two examples of successive catalyses by a single muon (Fig. 11). We also observed the catalysis of $D + D \rightarrow {}^3H + {}^1H$ in pure liquid deuterium.

A few months after we had announced our μ-catalysis results, the world of particle physics was shaken by the discovery that parity was not conserved in β-decay. Madame Wu and her collaborators[52], acting on a suggestion by Lee and Yang[53], showed that the β-rays from the decay of oriented ^{60}Co nuclei were emitted preferentially in a direction opposite to that of the spin. Lee and Yang suggested that parity nonconservation might also manifest itself in the weak decay of the Λ hyperon into a proton plus a negative pion. Crawford *et al.* had moved the 10-inch chamber into a negative pion beam, and were analyzing a large sample of Λ's from associated production events. They looked for an «up-down asymmetry» in the emission of pions from Λ's, relative to the «normal to the production plane», as suggested by Lee and Yang. As a

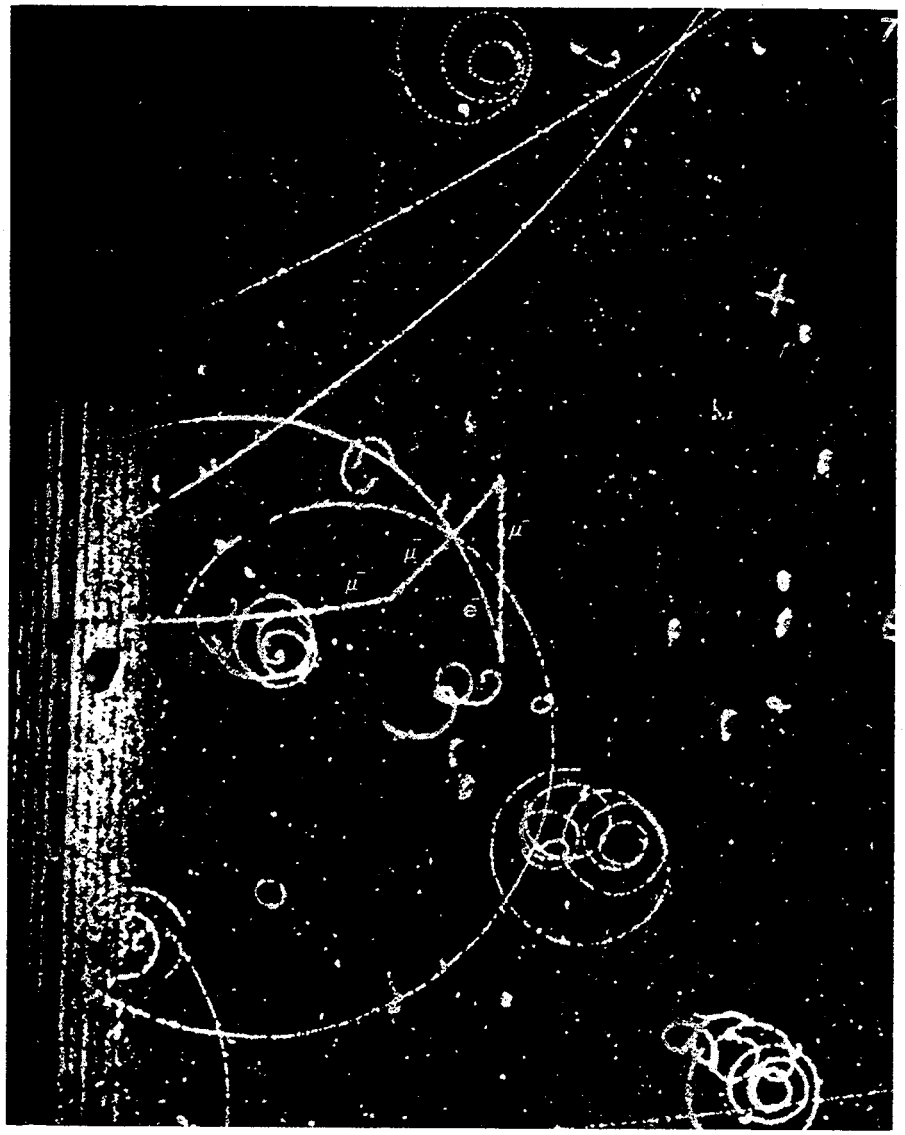


Fig. 11. Double muon catalysis.

result, they had the pleasure of being the first to observe parity nonconservation in the decay of hyperons[54].

In the winter of 1958, the 15-inch chamber had completed its engineering test run as a prototype for the 72-inch chamber, and was operating for the first time as a physics instrument. Harold Ticho, Bud Good and Philippe

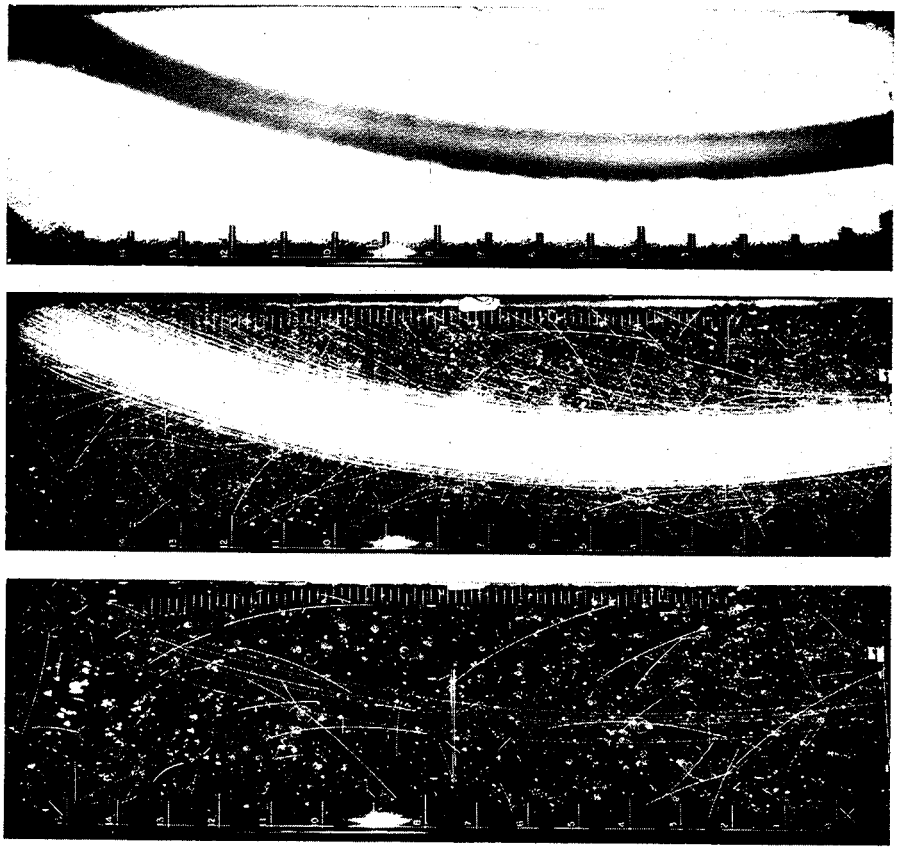


Fig.12. K^- beam in 72-inch bubble chamber. (a) No spectrometers on; (b) one spectrometer on; (c) two spectrometers on.

Eberhard[55] had designed and built the first separated beam of K^- mesons with a momentum of more than 1 GeV/c. Fig.12 shows the appearance of a bubble chamber when such a beam is passed through it, and when one or both of the electrostatic separators are turned off. The ingenuity which has been brought to bear on the problem of beam separation, largely by Ticho and Murray, is difficult to imagine, and its importance to the success of our program cannot be overestimated[55]. Joe Murray has recently joined the Stanford Linear Accelerator Center, where he has in a short period of time built a very successful radiofrequency-separated K beam and a backscattered laser beam.

The first problem we attacked with the 15-inch chamber was that of the Ξ^0. Gell-Mann had predicted that the Ξ^- was one member of an I-spin doublet, with strangeness minus 2. The predicted partner of the Ξ^- would be a

neutral hyperon that decayed into a Λ and a π^0 – both neutral particles that would, like the Ξ^0, leave no track in the bubble chamber. A few years earlier, as an after-dinner speaker at a physics conference, Victor Weisskopf had «brought down the house» by exhibiting an absolutely blank cloud-chamber photograph, and saying that it represented proof of the decay of a new neutral particle into two other neutral particles! And now we were seriously planning to do what had been considered patently ridiculous only a few years earlier.

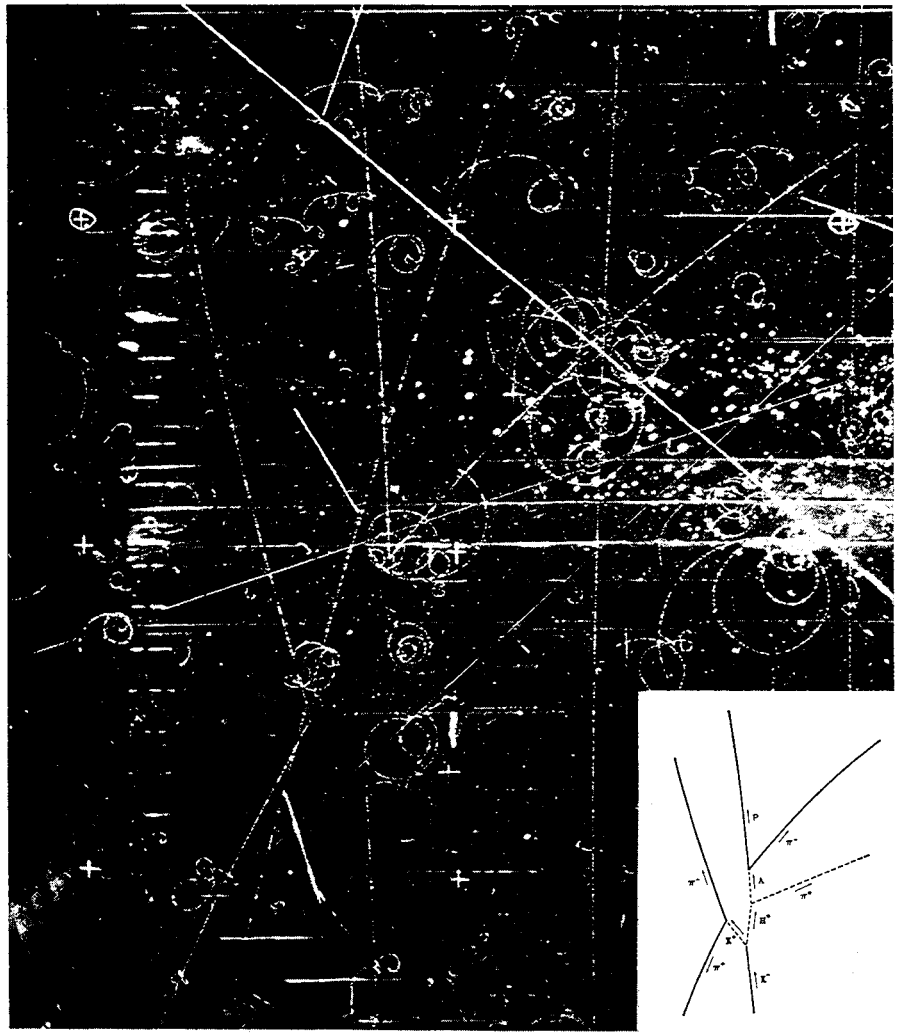


Fig. 13. Production and decay of a neutral cascade hyperon (Xi zero).

According to the Gell-Mann and Nishijima strangeness rules, the Ξ^0 should be seen in the reaction

$$\begin{array}{ccccc} K^- + p \rightarrow \Xi^0 & & + & & K^0 \\ & \downarrow & & & \downarrow \\ & \Lambda + \pi^0 & & & \pi^- + \pi^+ \\ & \downarrow & & & \\ & \pi^- + p & & & \end{array}$$

In the one example of this reaction that we observed, Fig.13, the charged pions from the decay of the neutral K^0 yielded a measurement of the energy and direction of the unobserved K^0. Through the conservation laws of energy and momentum (plus a measurement of the momentum of the interacting K^- track) we could calculate the mass of the coproduced Ξ^0 hyperon plus its velocity and direction of motion. Similarly, measurements of the π^- and proton gave the energy and direction of motion of the unobserved Λ, and proved that it did not come directly from the point at which the K^- meson interacted with the proton. The calculated flight path of the Λ intersected the calculated flight path of the Ξ^0, and the angle of intersection of the two unobserved but calculated tracks gave a confirming measurement of the mass of the Ξ^0 hyperon, and proved that it decayed into a Λ plus a π^0. This single hard-won event was a sort of tour de force that demonstrated clearly the power of the liquid hydrogen bubble chamber plus its associated data-analysis techniques.

Although only one Ξ^0 was observed in the short time the 15-inch chamber was in the separated K^- beam, large numbers of events showing strange-particle production were available for study. The Franckensteins were kept busy around the clock measuring these events, and those of us who had helped to build and maintain the beam now concentrated our attention on the analysis of these reactions. The most copious of the simple «topologies» was $K^-p\rightarrow$ two charged prongs plus a neutral V-particle. According to the strangeness rules, this topology could represent either

$$\begin{array}{c} K^- + p \rightarrow \Lambda + \pi^+ + \pi^- \\ \downarrow \\ \pi^- + p \end{array}$$

or

$$\begin{array}{c} K^- + p \rightarrow \overline{K}^0 + p + \pi^- \\ \downarrow \\ \pi^- + \pi^+ \end{array}$$

The kinematics program KICK was now available to distinguish between these two reactions, and to eliminate those examples of the same topology in

which an unobserved π^0 was produced at the first vertex. SUMX had not yet been written, so the labor of plotting histograms was assumed by the two very able graduate students who had been associated with the K^- beam and its exposure to the 15-inch chamber since its planning stages: Stanley Wojcicki and Bill Graziano. They first concentrated their attention on the energies of the charged pions from the production vertex in the first of the two reactions listed above. Since there were three particles produced at the vertex–a charged pion of each sign plus a Λ–one expected to find the energies of each of the three particles distributed in a smooth and calculable way from a minimum value to a maximum value. The calculated curve is known in particle physics as the «phase-space distribution». The decay of a τ meson into three charged pions was a well known «three-particle reaction» in which the dictates of phase space were rather precisely followed.

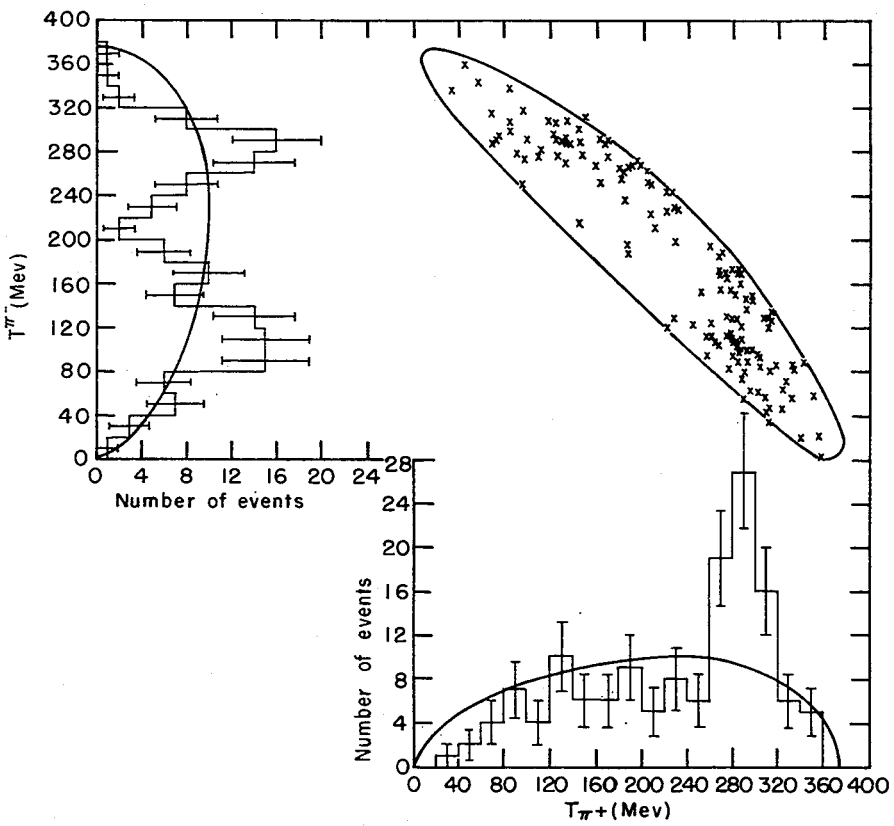


Fig. 14. Discovery of the Y_1^* (1385).

But when Wojcicki and Graziano finished transcribing their data from KICK printout into histograms, they found that phase-space distributions were poor approximations to what they observed. Fig. 14 shows the distribution of energy of both positive and negative mesons, together with the corresponding «Dalitz plot», which Richard Dalitz[56] had originated to elucidate the «τ–θ puzzle», which had in turn led to Lee and Yang's parity-nonconservation hypothesis.

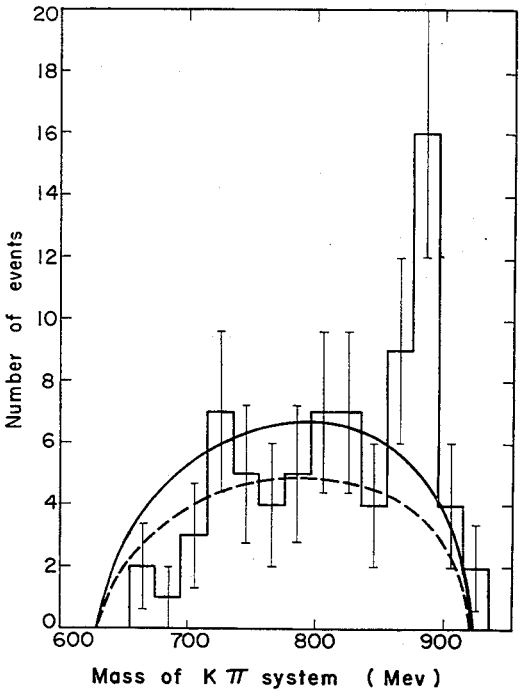


Fig. 15. Discovery of the K^* (890).

The peaked departure from a phase-space distribution had been observed only once before in particle physics, where it had distinguished the reaction $p+p\rightarrow\pi^+ + d$ from the «three-body reaction» $p+p\rightarrow\pi^+ + p + n$. (Although no new particles were discovered in these reactions, they did contribute to our knowledge of the spin of the pion[57].) But such a peaking had been observed in the earliest days of experimentation in the artificial disintegration of nuclei, and its explanation was known from that time. Oliphant and Rutherford[58] observed the reaction $p + {}^{11}\mathrm{B} \rightarrow 3\ {}^{4}\mathrm{He}$. This is a three-body reaction, and the energies of the α particles had a phase-space-like distribution except for the

fact that there was a sharp spike in the energy distribution at the highest α-particle energy. This was quickly and properly attributed[58] to the reaction

$$p + {}^{11}\mathrm{B} \rightarrow {}^{8}\mathrm{Be} + {}^{4}\mathrm{He}$$
$$\downarrow$$
$${}^{4}\mathrm{He} + {}^{4}\mathrm{He}$$

In other words, some of the reactions proceeded *via* a two-body reaction, in which one α particle recoiled with unique energy against a quasistable ^{8}Be nucleus. But the ^{8}Be nucleus was itself unstable, coming apart in 10^{-16} sec into two α particles of low relative energy. The proof of the fleeting existence of ^{8}Be was the peak in the high energy α-particle distribution, showing that initially only two particles, ^{8}Be and ^{4}He, participated in the reaction.

The peaks seen in Fig. 14 were thus a proof that the $\pi^{\pm}$ recoiled against a combination of $\Lambda + \pi^{\mp}$ that had a unique mass, broadened by the effects of the uncertainty principle. The mass of the $\Lambda\pi$ combination was easily calculable as 1385 MeV, and the I-spin of the system was obviously 1, since the I-spin of the Λ is 0, and the I-spin of the π is 1. This was then the discovery of the first «strange resonance», the $Y_1^*(1385)$. Although the famous Fermi 3,3-resonance had been known for years, and although other resonances in the π^- nucleon system had since shown up in total cross-section experiments at Brookhaven and Berkeley, CalTech and Cornell[59], the impact of the Y_1^* resonance on the thinking of particle physicists was quite different–*the* Y_1^* *really acted like a new particle*, and not simply as a resonance in a cross section.

We announced the Y_1^* at the 1960 Rochester High Energy Physics Conference[60], and the hunt for more short-lived particles began in earnest. The same team from our bubble-chamber group that had found the $Y_1^*(1385)$ now found two other strange resonances before the end of 1960–the $K^*(890)$[61], and the $Y_0^*(1405)$[62].

Although the authors of these three papers have for years been referred to as «Alston *et al.*», I think that on this occasion it is proper that the full list be named explicitly. In addition to Margaret Alston (now Margaret Garnjost) and Luis W. Alvarez, and still in alphabetical order, the authors are: Philippe Eberhard, Myron L. Good, William Graziano, Harold K. Ticho, and Stanley G. Wojcicki.

Figs. 15 and 16 show the histograms from the papers announcing these two new particles; the K^* was the first example of a «boson resonance» found by any technique. Instead of plotting these histograms against the energy of one particle, we introduced the now universally accepted technique of plotting

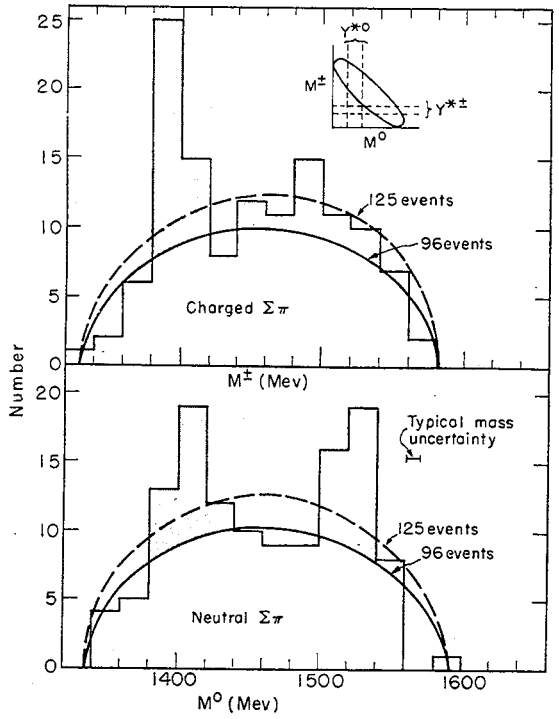


Fig. 16. Discovery of the Y_0^* (1405).

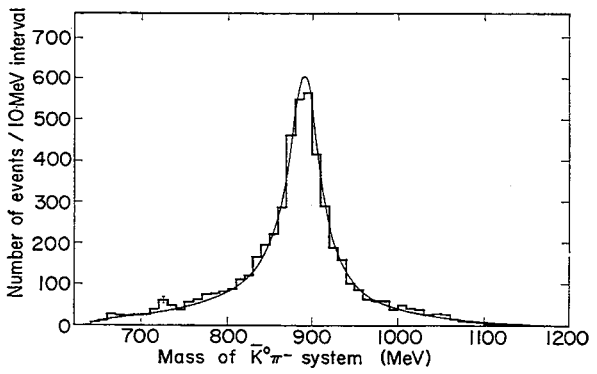


Fig. 17. Present-day K^* (890).

them against the effective mass of the composite system: $\Sigma+\pi$ for the Y_0^* (1405) and $K+\pi$ for the $K^*(890)$. Fig. 17 shows the present state of the art relative to the $K^*(890)$; there is essentially no phase–space background in this

histogram, and the width of the resonance is clearly measurable to give the lifetime of the resonant state *via* the uncertainty principle.

These three earliest examples of strange-particle resonances all had lifetimes of the order of 10^{-23} sec, so the particles all decayed before they could traverse more than a few nuclear radii. No one had foreseen that the bubble chamber could be used to investigate particles with such short lives; our chambers had been designed to investigate the strange particle swith lifetimes of 10^{-10} sec–10^{13} times as long.

In the summer of 1959, the 72-inch chamber was used in its first planned physics experiment. Lynn Stevenson and Philippe Eberhard designed and constructed a separated beam of about 1.6-GeV/c antiprotons, and a quick scan of the pictures showed the now famous first example of antilambda production, *via* the reaction

$$\begin{array}{ccccc} \bar{p} + p \rightarrow \bar{\Lambda} & + & \Lambda \\ \downarrow & & \downarrow \\ \pi^+ + \bar{p} & & \pi^- + p \end{array}$$

Fig. 18 shows this photograph, with the antiproton from the antilambda decay annihilating in a four-pion event. I believe that everyone who attended the 1959 High Energy Physics Conference in Kiev will remember the showing of this photograph–the first interesting event from the newly operating 72-inch chamber.

Hofstadter's classic experiments on the scattering of high-energy electrons by protons and neutrons[63] showed for the first time how the electric charge was distributed throughout the nucleons. The theoretical interpretation of the experimental results[64] required the existence of two new particles, the vector mesons now known as the ω and the ϱ. The adjective «vector» simply means that these two mesons have one unit of spin, rather than zero, as the ordinary π and K mesons have. The ω was postulated to have I-spin = 0, and the ϱ to have I-spin = 1; the ω would therefore exist only in the neutral state, while the ϱ would occur in the +, −, and 0 charged states.

Many experimentalists, using a number of techniques, set out to find these important particles, whose masses were only roughly predicted. The first success came to Bogdan Maglić, a visitor to our group, who analyzed film from the 72-inch chamber's antiproton exposure. He made the important decision to concentrate his attention on proton–antiproton annihilations into five pions–two negative, two positive, and one neutral. KICK gave him a selected sample of such events; the tracks of the π^0 couldn't be seen, of course, but the constraints of the conservation laws permitted its energy and direction

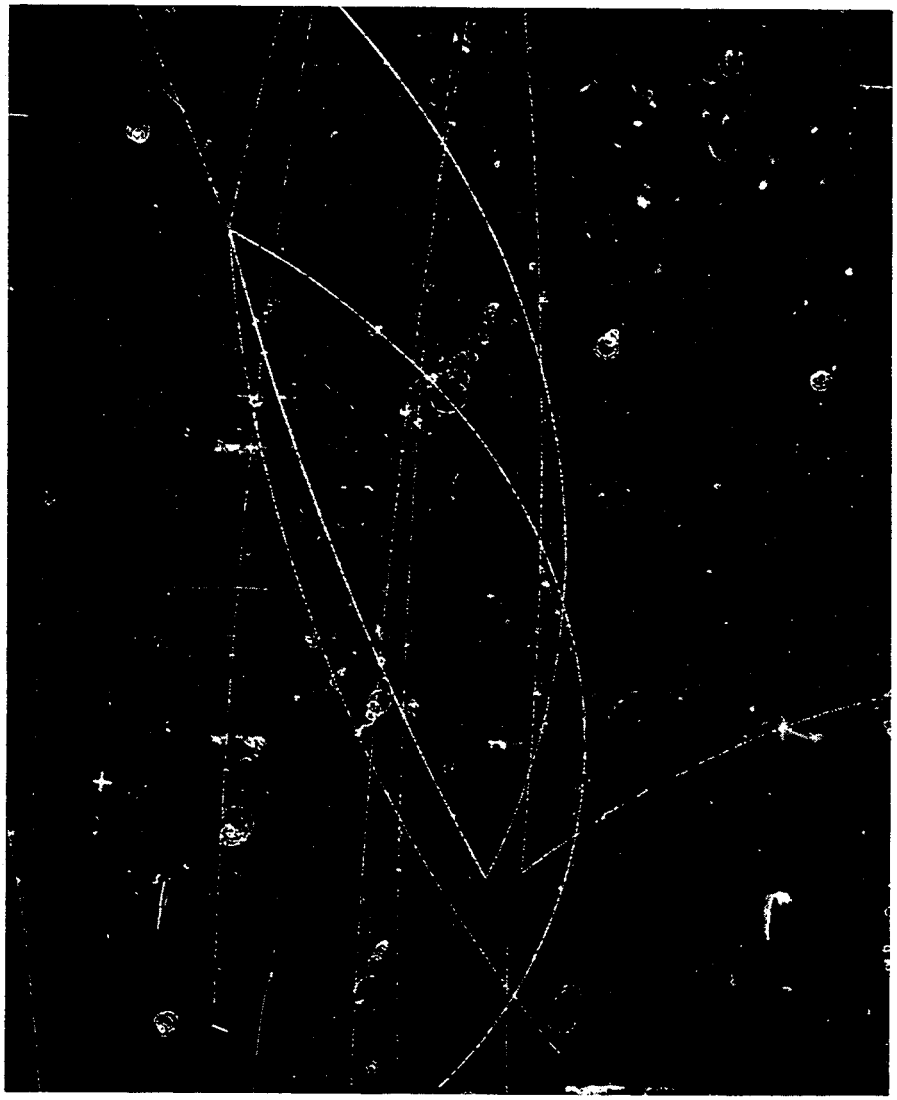


Fig. 18. First production of anti-lambda.

to be computed. Maglić then plotted a histogram of the effective mass of all neutral three-pion combinations. There were four such neutral combinations for each event; the neutral pion was taken each time together with all four possible pairs of oppositely charged pions. SUMX was just beginning to work, and still had bugs in it, so the preparation of the histogram was a very

tedious and time-consuming chore, but as it slowly emerged, Maglić had the thrill of seeing a bump appear in the side of his phase-space distribution. Fig. 19 shows a small portion of the whole distributions, with the peak that signaled the discovery of the very important ω meson.

Although Bogdan Maglić originated the plan for this search, and pushed through the measurements by himself, he graciously insisted that the paper announcing his discovery[65] should be co-authored by three of us who had developed the chamber, the beam, and the analysis program that made it possible.

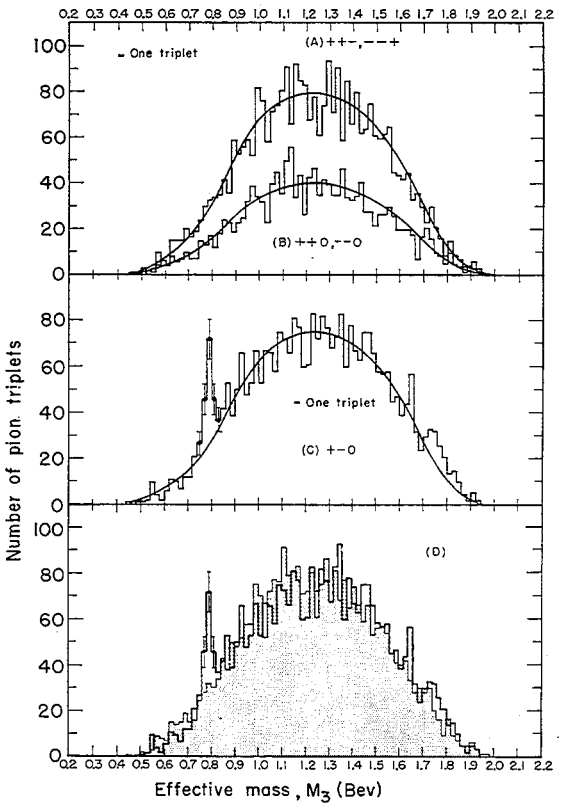


Fig. 19. Discovery of the ω meson.

The ϱ meson is the only one from this exciting period in the development of particle physics whose discovery cannot be assigned uniquely. In our group, the two Franckensteins were being used full time on problems that the senior members felt had higher priority. But a team of junior physicists and graduate

students, Anderson *et al.*[66], found that they could make accurate enough measurements directly on the scanning tables to accomplish a «Chew–Low extrapolation». Chew and Low had described a rather complicated procedure to look for the predicted dipion resonance now known as the ϱ meson. Fig. 20 shows the results of this work, which convinced me that the ϱ existed and had its predicted spin of 1. The mass of the ϱ was given as about 650 MeV, rather than its now accepted value of 765 MeV. (This low value is now explained in terms of the extreme width of the ϱ resonance.) The evidence for the ϱ seemed to me even more convincing than the early evidence Fermi and his co-workers produced in favor of the famous 3,3 pion–nucleon resonance.

But one of the unwritten laws of physics is that one really hasn't made a discovery until he has convinced his peers that he has done so. We had just persuaded high-energy physicists that the way to find new particles was to look for bumps on effective-mass histograms, and some of them were therefore unimpressed by the Chew–Low demonstration of the ϱ. Fortunately, Walker and his collaborators[67] at Wisconsin soon produced an effective-mass ideogram with a convincing bump at 765 MeV, and they are therefore most often listed as the discoverers of the ϱ.

Ernest Lawrence very early established the tradition that his laboratory would share its resources with others outside its walls. He supplied short-lived

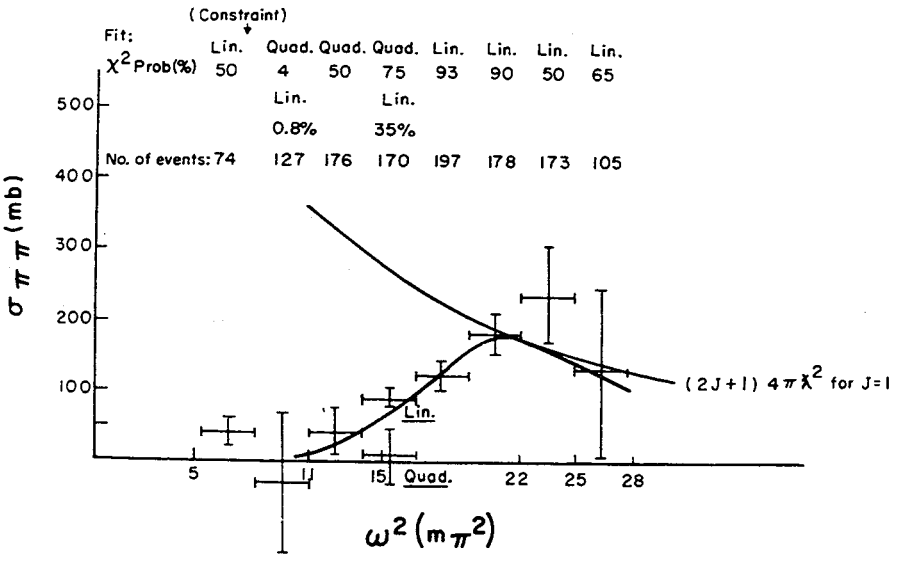


Fig. 20. First evidence for the ϱ meson.

radioactive materials to scientists in all departments at Berkeley, and he sent longer-lived samples to laboratories throughout the world. The first artificially created element, technetium, was found by Perrier and Segrè[68], who did their work in Palermo, Sicily. They analyzed the radioactivity in a molybdenum deflector strip from the Berkeley 28-inch cyclotron that had been bombarded for many months by 6-MeV deuterons.

We followed Ernest Lawrence's example, and thus participated vicariously in a number of important discoveries of new particles. The first was the η found at Johns Hopkins, by a group headed by Aihud Pevsner[69]. They analyzed film from the 72-inch chamber, and found the η with a mass of 550 MeV, decaying into $\pi^+\pi^-\pi^0$. Within a few weeks of the discovery of the η,

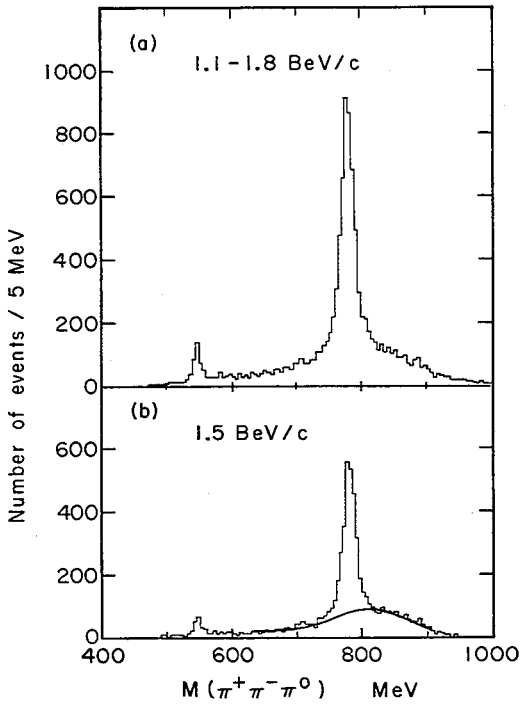


Fig. 21. Present-day histogram showing ω and η mesons.

Rosenfeld and his co-workers[70] at Berkeley, who had independently observed the η, showed quite unexpectedly that I spin was not conserved in its decay. Fig. 21 shows the present state of the art with respect to the ω and η mesons; the strengths of their signatures in this single histogram is in marked

contrast to their first appearances in 72-inch bubble-chamber experiments.

In the short interval of time between the first and second publications on the η, the discovery of the $Y_0^*(1520)$ was announced by Ferro-Luzzi, Tripp, and Watson[71], using a new and elegant method. Bob Tripp has continued to be a leader in the application of powerful methods of analysis to the study of the new particles.

The discovery of the $\Xi^*(1530)$ hyperon was accomplished in Los Angeles by Ticho and his associates[72], using 72-inch bubble-chamber film. Harold Ticho had spent most of his time in Berkeley for several years, working tirelessly on every phase of our work, and many of his colleagues had helped prepare the high-energy separated K^- beam for what came to be known as the K72 experiment. The UCLA group analyzed the two highest-momentum K^- exposures in the 72-inch chamber, and found the $\Xi^*(1530)$ just in time to report it at the 1962 High Energy Physics Conference in Geneva. (Confirming evidence for this resonance soon came from Brookhaven[73].)

Murray Gell-Mann had recently enunciated his important ideas concerning the «Eightfold Way»[74], but his paper had not generated the interest it deserved. It was soon learned that Ne'eman had published the same suggestions, independently[75].

The announcement of the $\Xi^*(1530)$ fitted exactly with their predictions of the mass and other properties of that particle. One of their suggestions was that four I-spin multiplets, all with the same spin and parity, would exist in a «decuplet» with a mass spectrum of «lines» showing an equal spacing. They put the Fermi 3,3-resonance as the lowest mass member, at 1238 MeV. The second member was the $Y_1^*(1385)$, so the third member should have a mass of $(1385)+(1385-1238)=1532$. The strangeness and the multiplicity of each member of the spectrum was predicted to drop 1 unit per member, so the $\Xi^*(1530)$ fitted their predictions completely. It was then a matter of simple arithmetic to set the mass, the strangeness, and the charge of the final member -the Ω^-. The realization that there was now a workable theory in particle physics was probably the high point of the 1962 International Conference on High Energy Physics.

Since the second and third members of the series-the ones that permitted the prediction of the properties of the Ω^- to be made-had come out of our bubble chambers, it was a matter of great disappointment to us that the Bevatron energy was insufficient to permit us to look for the Ω^-. Its widely acclaimed discovery[76] had to wait almost two years, until the 80-inch chamber at Brookhaven came into operation.

Since the name of the Ω had been picked to indicate that it was the last of the particles, the mention of its discovery is a logical point at which to conclude this lecture. I will do so, but not because the discovery of the Ω signaled the end of what is sometimes called the population explosion in particle physics – the latest list[77] contains between 70 and 100 particle multiplets, depending upon the degree of certainty one demands before «certification». My reason for stopping at this point is simply that I have discussed most of the particles found by 1962 – the ones that were used by Gell-Mann and Ne'eman to formulate their SU(3) theories – and things became much too involved after that time. So many groups were then in the «bump-hunting business» that most discoveries of new resonances were made simultaneously in two or more laboratories.

I am sorry that I have neither the time nor the ability to tell you of the great beauty and the power that has been brought to particle physics by our theoretical friends. But I hope that before long, you will hear it directly from them.

In conclusion, I would like to apologize to those of my colleagues and my friends in other laboratories, whose important work could not be mentioned because of time limitations. By making my published lecture longer than the oral presentation, I have reduced the number of apologies that are necessary, but unfortunately I could not completely eliminate such debts.

1. J. Chadwick, *Proc. Roy. Soc.* (*London*), *Ser. A*, 136 (1932) 692.
2. C. D. Anderson, *Science*, 76 (1932) 238.
3. M. Conversi, E. Pancini and O. Piccioni, *Phys. Rev.*, 71 (1947) 209.
4. S. H. Neddermeyer and C. D. Anderson, *Phys. Rev.*, 51 (1937) 884.
5. J. C. Street and E. C. Stevenson, *Phys. Rev.*, 51 (1937) 1005.
6. H. Yukawa, *Proc. Phys.-Math. Soc. Japan*, 17 (1935) 48.
7. C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini and C. F. Powell, *Nature*, 159 (1947) 694.
8. E. Gardner and C. M. G. Lattes, *Science*, 107 (1948) 270.
9. R. Bjorklund, W. E. Crandall, B. J. Moyer and H. F. York, *Phys. Rev.*, 77 (1950) 213.
10. J. Steinberger, W. K. H. Panofsky and J. Steller, *Phys. Rev.*, 78 (1950) 802.
11. A. G. Carlson (now A. G. Ekspong), J. E. Hooper and D. T. King, *Phil. Mag.*, 41 (1950) 701.
12. H. L. Anderson, E. Fermi, E. A. Long, R. Martin and D. E. Nagle, *Phys. Rev.*, 85 (1952) 934.
13. B. Cassen and E. U. Condon, *Phys. Rev.*, 50 (1936) 846.

14. K. A. Brueckner, *Phys. Rev.*, 86 (1952) 106.
15. W. Pauli and S. M. Dancoff, *Phys. Rev.*, 62 (1942) 85.
16. A. Pais, *Phys. Rev.*, 86 (1952) 663.
17. W. B. Fowler, R. P. Shutt, A. M. Thorndike and W. L. Whittemore, *Phys. Rev.*, 91 (1953) 1287; 93 (1954) 861; 98 (1955) 121.
18. O. Chamberlain, E. Segrè, C. Wiegand and T. Ypsilantis, *Phys. Rev.*, 100 (1955) 947.
19. G. D. Rochester and C. C. Butler, *Nature*, 160 (1947) 855.
20. A. J. Seriff, R. B. Leighton, C. Hsiao, E. D. Cowan and C. D. Anderson, *Phys. Rev.*, 78 (1950) 290.
21. R. Armenteros, K. H. Barker, C. C. Butler, A. Cachon and C. M. York, *Phil. Mag.*, 43 (1952) 597.
22. L. Leprince-Ringuet and M. l'Héritier, *Compt. Rend.*, 219 (1944) 618.
23. H. A. Bethe, *Phys. Rev.*, 70 (1946) 821.
24. R. M. Brown, U. Camerini, P. H. Fowler, H. Muirhead, C. F. Powell and D. M. Ritson, *Nature*, 163 (1949) 47.
25. C. O'Ceallaigh, *Phil. Mag.*, 42 (1951) 1032.
26. M. G. K. Menon and C. O'Ceallaigh, *Proc. Roy. Soc. (London), Ser. A*, 221 (1954) 292.
27. R. W. Thompson, A. V. Buskirk, L. R. Etter, C. J. Karzmark and R. H. Rediker, *Phys. Rev.*, 90 (1953) 329.
28. A. Bonetti, R. Levi Setti, M. Panetti and G. Tomasini, *Nuovo Cimento*, 10 (1953) 345.
29. R. Marshak and H. Bethe, *Phys. Rev.*, 72 (1947) 506.
30. M. Gell-Mann, *Phys. Rev.*, 92 (1953) 833.
31. K. Nishijima, *Progr. Theoret. Phys. (Kyoto)*, 12 (1954) 107.
32. L. W. Alvarez and S. Goldhaber, *Nuovo Cimento*, 2 (1955) 344.
33. T. D. Lee, Weak interactions and nonconservation of parity, in *Nobel Lectures Physics, 1942–1962*, Elsevier, Amsterdam, 1964, p. 406.
34. C. N. Yang, The law of parity conservation and other symmetry laws of physics, in *Nobel Lectures Physics, 1942–1962*, Elsevier, Amsterdam, 1964, p. 393.
35. L. W. Alvarez, F. S. Crawford Jr., M. L. Good and M. L. Stevenson, *Phys. Rev.*, 101 (1956) 303.
36. V. Fitch and R. Motley, *Phys. Rev.*, 101 (1956) 496.
37. S. von Friesen, *Arkiv Fysik*, 8 (1954) 309; 10 (1956) 460.
38. R. W. Birge, D. H. Perkins, J. R. Peterson, D. H. Stork and M. N. Whitehead, *Nuovo Cimento*, 4 (1956) 834.
39. D. A. Glaser, Elementary particles and bubble chambers, in *Nobel Lecture Physics, 1942–1962*, Elsevier, Amsterdam, 1964, p. 529.
40. R. H. Hildebrand and D. E. Nagle, *Phys. Rev.*, 92 (1953) 517.
41. J. G. Wood, *Phys. Rev.*, 94 (1954) 731.
42. D. P. Parmentier and A. J. Schwemin, *Rev. Sci. Instr.*, 26 (1955) 958.
43. D. C. Gates, R. W. Kenney and W. P. Swanson, *Phys. Rev.*, 125 (1962) 1310.
44. L. W. Alvarez, F. S. Crawford Jr. and M. L. Stevenson, *Phys. Rev.*, 112 (1958) 1267.
45. L. W. Alvarez, P. Davey, R. Hulsizer, J. Snyder, A. J. Schwemin and R. Zane, *UCRL -10109*, 1962 (unpublished); P. G. Davey, R. I. Hulsizer, W. E. Humphrey, J. H. Munson, R. R. Ross and A. J. Schwemin, *Rev. Sci. Instr.*, 35 (1964) 1134.

46. L. W. Alvarez, in *Proceedings of the 1966 International Conference on Instrumentation for High Energy Physics*, Stanford, Calif., p. 271.
47. W. K. H. Panofsky, L. Aamodt and H. F. York, *Phys. Rev.*, 78 (1950) 825.
48. L. W. Alvarez, H. Bradner, P. Falk-Vairant, J. D. Gow, A. H. Rosenfeld, F. T. Solmitz and R. D. Tripp, *Nuovo Cimento*, 5 (1957) 1026.
49. L. W. Alvarez, H. Bradner, F. S. Crawford Jr., J. A. Crawford, P. Falk-Vairant, M. L. Good, J. D. Gow, A. H. Rosenfeld, F. T. Solmitz, M. L. Stevenson, H. K. Ticho and R. D. Tripp, *Phys. Rev.*, 105 (1957) 1127.
50. F. C. Frank, *Nature*, 160 (1947) 525.
51. Ya. B. Zel'dovitch, *Dokl. Akad. Nauk S. S. S. R.*, 95 (1954) 493.
52. C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Phys. Rev.*, 105 (1957) 1413.
53. T. D. Lee and C. N. Yang, *Phys. Rev.*, 104 (1956) 254, 822.
54. F. S. Crawford Jr., M. Cresti, M. L. Good, K. Gottstein, E. M. Lyman, F. T. Solmitz, M. L. Stevenson and H. K. Ticho, *Phys. Rev.*, 108 (1957) 1102.
55. P. Eberhard, M. L. Good and H. Ticho, *UCRL-8878*, Aug. 1959 (unpublished); J. J. Murray, *UCRL-3492*, May 1957 (unpublished); J. J. Murray, *UCRL-9506*, Sept. 1960 (unpublished).
56. R. H. Dalitz, *Phil. Mag.*, 44 (1953) 1068.
57. W. F. Cartwright, C. Richman, M. N. Whitehead and H. A. Wilcox, *Phys. Rev.*, 78 (1950) 823; D. L. Clark, A. Roberts and R. Wilson, *Phys. Rev.*, 83 (1951) 649; R. Durbin, H. Loar and J. Steinberger, *Phys. Rev.*, 83 (1951) 646.
58. M. L. E. Oliphant and E. Rutherford, *Proc. Roy. Soc. (London)*, Ser. A, 141 (1933) 259; M. L. E. Oliphant, A. E. Kempton and E. Rutherford, *Proc. Roy. Soc. (London)*, Ser. A, 150 (1935) 241.
59. R. L. Cool, L. Madansky and O. Piccioni, *Phys. Rev.*, 93 (1954) 637; see also refs. in R. F. Peierls, *Phys. Rev.*, 118 (1959) 325.
60. M. Alston, L. W. Alvarez, P. Eberhard, M. L. Good, W. Graziano, H. K. Ticho and S. G. Wojcicki, *Phys. Rev. Letters*, 5 (1960) 520.
61. M. Alston, L. W. Alvarez, P. Eberhard, M. L. Good, W. Graziano, H. K. Ticho and S. G. Wojcicki, *Phys. Rev. Letters*, 6 (1961) 300.
62. M. Alston, L. W. Alvarez, P. Eberhard, M. L. Good, W. Graziano, H. K. Ticho and S. G. Wojcicki, *Phys. Rev. Letters*, 6 (1961) 698.
63. R. Hofstadter, *Rev. Mod. Phys.*, 28 (1956) 214.
64. W. Holladay, *Phys. Rev.*, 101 (1956) 1198; Y. Nambu, *Phys. Rev.*, 106 (1957) 1366; G. F. Chew, *Phys. Rev. Letters*, 4 (1960) 142; W. R. Frazer and J. R. Fulco, *Phys. Rev.*, 117 (1960) 1609; F. J. Bowcock, W. N. Cottingham and D. Lurie, *Phys. Rev. Letters*, 5 (1960) 386.
65. B. C. Maglić, L. W. Alvarez, A. H. Rosenfeld and M. L. Stevenson, *Phys. Rev. Letters*, 7 (1961) 178.
66. J. A. Anderson, V. X. Bang, P. G. Burke, D. D. Carmony and N. Schmitz, *Phys. Rev. Letters*, 6 (1961) 365.
67. A. R. Erwin, R. March, W. D. Walker and E. West, *Phys. Rev. Letters*, 6 (1961) 628.
68. C. Perrier and E. Segrè, *Atti Accad. Nazl. Lincei, Rend. Classe Sci. Fis. Mat. e Nat.*, 25 (1937) 723.

69. A. Pevsner, R. Kraemer, M. Nussbaum, C. Richardson, P. Schlein, R. Strand, T. Toohig, M. Block, A. Engler, R. Gessaroli and C. Meltzer, *Phys. Rev. Letters*, 7 (1961) 421.
70. P. L. Bastien, J. P. Berge, O. I. Dahl, M. Ferro-Luzzi, D. H. Miller, J. J. Murray, A. H. Rosenfeld and M. B. Watson, *Phys. Rev. Letters*, 8 (1962) 114.
71. M. Ferro-Luzzi, R. D. Tripp and M. B. Watson, *Phys. Rev. Letters*, 8 (1962) 28.
72. G. M. Pjerrou, D. J. Prowse, P. Schlein, W. E. Slater, D. H. Stork and H. K. Ticho, *Phys. Rev. Letters*, 9 (1962) 114.
73. L. Bertanza, V. Brisson, P. L. Connolly, E. L. Hart, I. S. Mittra, G. C. Moneti, R. R. Rau, N. P. Samios, I. O. Skillicorn, S. S. Yamamoto, M. Goldberg, L. Gray, J. Leitner, S. Lichtman and J. Westgard, *Phys. Rev. Letters*, 9 (1962) 180.
74. M. Gell-Mann, *California Institute of Technology Synchrotron Laboratory Report CTSL-20*, 1961 (unpublished).
75. Y. Ne'eman, *Nucl. Phys.*, 26 (1961) 222.
76. V. E. Barnes, P. L. Connolly, D. J. Crennell, B. B. Culwick, W. C. Delaney, W. B. Fowler, P. E. Hagerty, E. L. Hart, N. Horwitz, P. V. C. Hough, J. E. Jensen, J. K. Kopp, K. W. Lai, J. Leitner, J. L. Lloyd, G. W. London, T. W. Morris, Y. Oren, R. B. Palmer, A. G. Prodell, D. Radojičič, D. C. Rahm, C. R. Richardson, N. P. Samios, J. R. Sanford, R. P. Shutt, J. R. Smith, D. L. Stonehill, R. C. Strand, A. M. Thorndike, M. S. Webster, W. J. Willis and S. S. Yamamoto, *Phys. Rev. Letters*, 12 (1964) 204.
77. A. H. Rosenfeld, A. Barbaro-Galtieri, W. J. Podolsky, L. R. Price, P. Söding, C. G. Wohl, M. Roos and W. J. Willis, *Rev. Mod. Phys.*, 39 (1967) 1.

Biography

Luis W. Alvarez was born in San Francisco, Calif., on June 13, 1911. He received his B. Sc. from the University of Chicago in 1932, a M. Sc. in 1934, and his Ph. D. in 1936. Dr. Alvarez joined the Radiation Laboratory of the University of California, where he is now a professor, as a research fellow in 1936. He was on leave at the Radiation Laboratory of the Massachusetts Institute of Technology from 1940 to 1943, at the Metallurgical Laboratory of the University of Chicago in 1943–1944, and at the Los Alamos Laboratory of the Manhattan District from 1944 to 1945.

Early in his scientific career, Dr. Alvarez worked concurrently in the fields of optics and cosmic rays. He is co-discoverer of the «East–West effect» in cosmic rays. For several years he concentrated his work in the field of nuclear physics. In 1937 he gave the first experimental demonstration of the existence of the phenomenon of *K*-electron capture by nuclei. Another early development was a method for producing beams of very slow neutrons. This method subsequently led to a fundamental investigation of neutron scattering in ortho- and para-hydrogen, with Pitzer, and to the first measurement, with Bloch, of the magnetic moment of the neutron. With Wiens, he was responsible for the production of the first ^{198}Hg lamp; this device was developed by the Bureau of Standards into its present form as the universal standard of length. Just before the war, Alvarez and Cornog discovered the radioactivity of ^{3}H (tritium) and showed that ^{3}He was a stable constituent of ordinary helium. (Tritium is best known as a source of thermonuclear energy, and ^{3}He has become of importance in low temperature research.)

During the war (at M.I.T.) he was responsible for three important radar systems – the microwave early warning system, the Eagle high altitude bombing system, and a blind landing system of civilian as well as military value (GCA, or Ground-Controlled Approach). While at the Los Alamos Laboratory, Professor Alvarez developed the detonators for setting off the plutonium bomb. He flew as a scientific observer at both the Almagordo and Hiroshima explosions.

Dr. Alvarez is responsible for the design and construction of the Berkeley

40-foot proton linear accelerator, which was completed in 1947. In 1951 he published the first suggestion for charge exchange acceleration that quickly led to the development of the «Tandem Van de Graaf accelerator». Since that time, he has engaged in high-energy physics, using the 6 billion electron volt Bevatron at the University of California Radiation Laboratory. His main efforts have been concentrated on the development and use of large liquid hydrogen bubble chambers, and on the development of high-speed devices to measure and analyze the millions of photographs produced each year by the bubble-chamber complex. The net result of this work has been the discovery by Dr. Alvarez' research group, of a large number of previously unknown «fundamental particle resonances». Since 1967 Dr. Alvarez has devoted most of this time to the study of cosmic rays, using balloons and superconducting magnets.

Professor Alvarez is a member of the following societies: National Academy of Sciences, American Philosophical Society, American Physical Society (President 1969), American Academy of Arts and Sciences, and National Academy of Engineering. In 1946 he was awarded the Collier Trophy by the National Aeronautical Association for the development of Ground-Controlled Approach. In 1953 he was awarded the John Scott Medal and Prize, by the city of Philadelphia, for the same work. In 1947 he was awarded the Medal for Merit. In 1960 he was named «California Scientist of the Year» for his research work on high-energy physics. In 1961 he was awarded the Einstein Medal for his contribution to the physical sciences. In 1963 he was awarded the Pioneer Award of the AIEEE; in 1964 he was awarded the National Medal of Science for contributions to high-energy physics, and in 1965 he received the Michelson Award. He has received the following honorary degrees: Sc.D., University of Chicago, 1967; Sc.D., Carnegie-Mellon University, 1968; Sc.D., Kenyon College, 1969.

Physics 1969

MURRAY GELL-MANN

«for his contributions and discoveries concerning the classification of elementary particles and their interactions»

Physics 1969

Presentation Speech by Professor Ivar Waller, member of the Nobel Committee for Physics

Your Majesty, Your Royal Highnesses, Ladies and Gentlemen.

Elementary particle physics which is now so vigorous was still in its infancy when Murray Gell-Mann in 1953 published the first of the papers which have been honoured with this years Nobel prize in physics.

The physicists were, however, already then aquainted with a rather large number of particles which apparently were indivisible and therefore elementary building stones of all matter. The elementary particle known for the longest time was the electron.

New particles were added when the atomic nuclei were explored. It was found that the atomic nuclei consist of positively charged protons and electrically neutral neutrons. These particles are held together in the atomic nuclei by enormously strong forces called nuclear forces which do not distinguish between protons and neutrons. This symmetry of the nuclear forces was expressed by saying that the nuclear forces are charge-independent. A proton and a neutron have further very nearly the same mass. They form a doublet of particles and have been given the common name of nucleons.

An increase already expected and desired occurred in the family of elementary particles at the end of the 1940's, when new particles called pi-mesons were discovered. They were named mesons because they have a mass between the electron and the nucleon masses. The pi-mesons had been predicted by the Japanese physicist Yukawa. They form a triplet of particules having nearly the same mass but different charges which are $+1$, 0 and -1 in units of the proton charge. Their interaction with the nucleons is strong and charge-independent. Their most important task is to be an intermediary agent for the strong interactions between the nucleons.

A very remarkable discovery which marked a new area in particle physics was made by the British physicists Rochester and Butler about the same time. They found new unstable particles which did not fit in with the theoretical ideas developed so far. Some of the new particles are heavier than the nucleons and were grouped together with them under the common name of baryons. The others were lighter than the nucleons but heavier than the electrons and

were called *K*-mesons. The new particles were copiously produced when high-energy pi-mesons collide with nucleons and were therefore assumed to interact strongly with other particles. But they had such a long lifetime that some law must exist which prevent the strong forces to act when they disintegrate into other particles. Gell-Mann discovered this law after some preliminary results had been found by Pais.

It had been assumed earlier that the new baryons from doublets like the nucleons and that the *K*-mesons form triplets like the pi-mesons. Gell-Mann made the fundamental new assumption that the new baryons instead form a singlet, a triplet and a doublet, the latter being different from the nucleon doublet, and that the new mesons form two kinds of doublets, one consisting of the antiparticles of the other. Gell-Mann assumed further that the principle of charge-independence was generally valid for strong interactions. He could thereby explain the mysterious properties of the new particles. He introduced a new fundamental characteristic of a multiplet called its hypercharge. This is defined as twice the mean value of the charges in the multiplet. Gell-Mann's proposed the new rule: Elementary particles can be transformed in others by the strong and the electromagnetic interactions only if the total hypercharge is conserved. This rule reminds of the law of conservation of the electric charge. It should be remarked that Gell-Mann initially used instead of the hypercharge a closely related number called the strangeness.

This discovery by Gell-Mann was admirable considering in particular the very meagre experimental material available to him. In the predicted baryon multiplets there occurred empty places. Gell-Mann could on this ground predict two new baryons. One of them was soon discovered but the other not until six years later.

This classification of the elementary particles and their interaction discovered by Gell-Mann has turned out to applicable to all strongly interacting particles found later and these are practically all particles discovered after 1953. His discovery is therefore fundamental in elementary particle physics.

It should be added that two Japanese physicists, Nakano and Nishijima, published a similar classification some months later than Gell-Mann.

Many theoretical physicists tried during the following years to find new symmetries which should give relations between the particle multiplets. Initiated by Sakata a series of papers were published in particular by Japanese physicists. They indicated that a certain kind of symmetry could be of interest. Gell-Mann showed in a new fundamentally important paper of 1961 that this symmetry which had since long been studied in pure mathematics could be

used for the classification of all strongly interacting particles. Assuming the validity of the new symmetry which includes the symmetry corresponding to charge-independence, Gell-Mann found that his earlier multiplets could be brought together into larger groups called supermultiplets each containing all baryons or all mesons which have the same spin and the same parity, *i.e.* have the same measure for their rotation around their axes and are transformed in the same way by reflections. Gell-Mann called this classification «The Eightfold Way». The nucleons were found to belong to a supermultiplet of eight particles *i.e.* an octet. For the mesons an octet was proposed were the pi- and *K*-mesons filled seven places. Because one place was empty a new meson was predicted. Its existence had been suspected already by some of the Japanese physicists mentioned above. It was soon discovered which meant that Gell-Mann's theory was strongly supported. Still more famous is Gell-Mann's prediction in 1962 of a new baryon called omega minus.

A similar classification was proposed by Y. Ne'eman somewhat later than Gell-Mann.

Gell-Mann has also found that «The Eightfold Way» can be described very simply by assuming that all particles which interact strongly with each other are composed of only three kinds of particles which he called quarks and of the corresponding antiparticles. The quarks are peculiar in particular because their charges are fractions of the proton charge which according to all experience up to now is the indivisible elementary charge. It has not yet been possible to find individual quarks although they have been eagerly looked for. Gell-Mann's idea is none the less of great heuristic value.

And interesting application of «The Eightfold Way» is the so-called current algebra which was founded by Gell-Mann. It has *e.g.* made evident that there are important connections between the different kinds of elementary particle interactions.

Gell-Mann has given many fundamental contributions to the theory of elementary particles besides those which have been mentioned here. He has during more than a decade been considered as the leading scientist in this field.

Professor Gell-Mann. You have given fundamental contributions to our knowledge of mesons and baryons and their interactions. You have developed new algebraic methods which have led to a far-reaching classification of these particles according to their symmetry properties. The methods introduced by you are among the most powerful tools for further research in particle physics.

On behalf of the Royal Swedish Academy of Science, I congratulate you on your successful work and ask you to receive your Nobel Prize from the hands of His Majesty the King.

MURRAY GELL-MANN

Symmetry and currents in particle physics

Nobel Lecture, December 11, 1969

[Professor Gell-Mann has presented his Nobel Lecture, but did not submit a manuscript for inclusion in this volume]

Biography

Murray Gell-Mann was born on 15th September 1929, in New York City. He obtained his B. Sc. at Yale University in 1948, and his Ph. D. in 1951 at the Massachusetts Institute of Technology. In 1952 he became a member of the Institute for Advanced Study, during 1952–1953 he was instructor at the University of Chicago, from 1953 to 1954 he was Assistant Professor, in 1954 he was appointed Associate Professor for research on dispersion relations. In this period he developed the strangeness theory and the eightfold way theory. In 1956 he was appointed Professor, his research then turned more to the theory of weak interactions.

In 1959 Professor Gell-Mann was awarded the Dannie Heineman Prize of the American Physical Society. He is a Fellow of this society and a member of the National Academy of Sciences.

Murray Gell-Mann was in 1955 married to J. Margaret Dow; they have a daughter, Elizabeth, and a son, Nicholas.

Physics 1970

HANNES ALFVÉN

«for fundamental work and discoveries in magneto-hydrodynamics with fruitful applications in different parts of plasma physics»

LOUIS NÉEL

«for fundamental work and discoveries concerning antiferromagnetism and ferri-magnetism which have led to important applications in solid state physics»

Physics 1970

Presentation Speech by Professor Torsten Gustafsson, member of the Royal Swedish Academy of Sciences

Your Majesty, Your Royal Highnesses, Ladies and Gentlemen.

From the sun, there blows a wind so hot that its atoms are split into electrically-charged particles, electrons and ions. They are attracted by the earth's magnetic field and the electrons follow the lines of force and produce the aurora borealis. This wind is one example of a plasma, an electrically-conducting gas with such remarkable properties that one, in addition to the well-known states of matter, solid, liquid and gaseous, has now, in the last fifty years, recognised it as a fourth. It is the most common state of matter in the universe. It was the most important state at the time of the creation of the solar and planetary systems; it is found in interstellar space, in fusion reactors and in welding apparatus.

Alfvén introduced into discussion of the aurora the fundamental idea that plasma, even in space, has a magnetic field associated with it.

In this way, he was led to study the general question of the significance of magnetic fields in the movements of plasmas. The magnetic field forces the positive and negative charges to move in different directions, giving rise to electric currents. The interaction of these currents produces mechanical forces, which can completely change the plasma's direction and speed. In particular, Alfvén discovered the existence of hitherto unsuspected magneto-hydrodynamical waves, the so-called Alfvén waves.

In cosmic physics, Alfvén's fundamental contribution has been the introduction of the magnetic field of force and the application of magneto-hydrodynamics. Prior to his work, one simply did not take these forces into consideration: through him, they have found widespread application in astrophysical problems, particularly in studying that phase of the development of the solar system in which the planets and satellites were created. Thus, the sun's rotation and the regular pattern of the planetary orbits can be explained by the idea that hydro-magnetic waves from the sun flowed along magnetic lines of force and transferred rotational energy to the planets when they were in the early stages of formation.

Furthermore, magneto-hydrodynamics is important in discussing the

problem of how the central body in a plasma cloud can develop into a sun and system of planets, or in investigations of stability conditions for a plasma consisting of electrons and ions moving at relativistic velocities interacting with cosmic fields. This is of interest in connection with both supernovae and the powerful eruptions recently found to occur in the centre of the Milky Way.

Alfvén's contributions to clarify the physical properties of plasmas have been considerable. Particularly important have been those works which form the bases of fusion research in different parts of the world. These works are important independently of how a fusion reactor can be built. The problem of containing a plasma at temperatures of millions of degrees in a magnetic field is related to Alfvén's concept of frozen lines of magnetic force. The plasma flowing in the bottle must not collapse like a breaking wave. Knowledge of the properties of Alfvén waves has been of extreme assistance in finding currents with the stability required.

Professor Alfvén. You have created magneto-hydrodynamics. Its development, in which you have played the major role, has shown the significance of this new branch of physics, both on the cosmic scale as well as here on earth.

On behalf of the Royal Academy of Science, it is my pleasure to congratulate you on your Nobel Prize in Physics.

About two thousand years ago, the first magnetic compass was made in China by stroking a piece of iron with a lump of magnetite. Such a compass always arouses much surprise, from the child who asks about the invisible force which aligns it along the north-south axis, to the scientist, who here confronts one of the very difficult problems of physics.

Three states of magnetism have long been recognised, dia-, para- and ferromagnetism. In the two former, the elementary magnets of the atoms behave independently of one another when subjected to a magnetic field. However, in ferromagnetism, which is many times stronger, they are aligned collectively, which makes the understanding of the physics much more difficult.

The first scientist who tried to explain magnetism was Ampère with his hypothesis about elementary currents. In 1907, Pierre Weiss found that there must be a special kind of force which aligned the elementary magnets, although he could not identify it. In his doctor's thesis in 1911. Niels Bohr showed that magnetism could not be caused by currents originating from the classical motion of electrical charges, but that something completely new was needed. Using the new ideas of atomic physics, Heisenberg in 1928 was able to give

a qualitative explanation of the aligning force occurring in ferromagnetics.

To these three types of magnetism, Néel in 1932 added a fourth, anti-ferromagnetism. He found that for certain crystals adjacent elementary magnets could align themselves anti-parallel and not parallel as in ferro-magnetic materials. He deduced the existence of a new variant of the force postulated by Weiss and presented a model for crystals which are built up from two interlaced lattices with equally strong magnetic fields acting in opposite directions. Anti-ferromagnetism is an ordered state with important properties. Thus, Néel showed that the magnetic state should disappear at a temperature now known as the Néel point, in analogy with the Curie point. Similarly, other remarkable observations in the physics of the solid state were explained.

In 1948, Néel made another fundamental discovery with his explanation of the strong magnetism found in the ferrite materials, of which magnetite is one. He generalized his earlier assumption by assuming that the lattices could be of different strengths and could produce external fields. In magnetite, with three atoms of iron and four of oxygen, the effects of two of the iron atoms cancel out while the third gives rise to the magnetic field. It is remarkable that magnetite which in the hands of the Chinese was used to produce the first compass, is in fact not ferromagnetic, but, in Néel's terminology, ferrimagnetic.

Néel could present an accurate description of the behaviour of the new synthetic magnetic materials and so explain hitherto puzzling experimental observations. These developments have been of considerable technical importance, *e.g.* in computer memories and in high-frequency techniques.

Néel has made many other contributions, such as investigations in the theory of magnetic domains and the discovery of the effect found in small particles, called super-paramagnetism.

Professor Néel. I have attempted to describe your major discoveries which follow in the great French tradition of studies of magnetic phenomena.

I have particularly emphasised your discoveries of anti-ferro- and ferrimagnetism which have been of such importance in the shaping of modern theories of magnetism.

I have the pleasure and the honour to convey to you the most sincere congratulations of the Royal Academy of Science.

Professor Alfvén, Professor Néel. I invite you to receive the Nobel Prize in Physics from the hands of His Majesty the King.

HANNES ALFVÉN

Plasma physics, space research and the origin of the solar system

Nobel Lecture, December 11, 1970

1. Science and instruments

The center of gravity of the physical sciences is always moving. Every new discovery displaces the interest and the emphasis. Equally important is that new technological developments open new fields for scientific investigation. To a considerable extent the way science takes depends on the construction of new instruments as is evident from the history of science. For example after the development of classical mechanics and electromagnetism during the 19th century, a new era was started by the construction of highly developed spectrographs in the beginning of this century. For its time those were very complicated and expensive instruments. They made possible the exploration of the outer regions of the atom. Similarly, in the thirties the cyclotron – for its time a very complicated and expensive instrument – was of major importance in the exploration of the nucleus. Finally, the last decade has witnessed the construction of still more complicated and expensive instruments, the space vehicles, which are launched by a highly developed rocket technology and instrumented with the most sophisticated electronic devices. We may then ask the question: What new fields of research – if any – do these open for scientific investigation? Is it true, also in this case, that the center of gravity of physics moves with the big instruments?

2. Scientific aims of space research

The first decade of space research mainly concentrated on the exploration of space near the Earth: the magnetosphere and interplanetary space. These regions earlier were supposed to be void and structureless but we now know that they are filled with plasmas, intersected by sheathlike discontinuities, and permeated by a complicated pattern of electric currents and electric and magnetic fields. The knowledge gained in this way is fundamental to our

general understanding of plasmas, especially cosmic plasmas. Indirectly it will hence be important to thermonuclear research, to the study of the structure of the galaxy and the metagalaxy, and to cosmological problems. Our advancing knowledge in cosmical electrodynamics will make it possible to approach these fields in a less speculative way than hitherto. The knowledge of plasmas is also fundamental to our understanding of the origin and evolution of the Solar System, because there are good reasons to believe that the matter which now forms the celestial bodies once was dispersed in a plasma state.

The second decade of space research seems to display a different character, at least to some extent. As several of the basic problems of the magnetosphere and interplanetary space are still unsolved, one can be sure that these regions will still command much interest. However, the lunar landings and also the deep-space probes to Venus and Mars have supplied us with so many new scientific facts that the emphasis in space research is moving towards the exploration of the Moon, the planets, and other celestial bodies in the Solar System.

The first phase of this exploration is necessarily of a character somewhat similar to the exploration of the polar regions and other regions of the earth which have been difficult to reach: a detailed mapping-out combined with geological, seismic, magnetic, and gravity surveys and an exploration of the atmospheric conditions. However, when applying this research pattern to the Moon and the planets one is confronted with another problem, *viz.* how these bodies were originally formed. In fact many of the recent space research reports end with speculations about the formation and evolution of the solar system. It seems that this will necessarily be one of the main problems–perhaps the main problem–on which space research will center in the near future. Already at an early data NASA stated that the main scientific goal of space research should be to clarify how the solar system was formed. This is indeed one of the fundamental problems of science. We are trying to write the scientific version of how our Earth and its neighbors once were created. From a – shall we say – philosophical point of view, this is just as important as the structure of matter, which has absorbed most of the interest during the first two thirds of this century.

3. Plasma physics and its applications

Before we concentrate on our main topic: how the solar system originated, we should make a brief summary of the state of plasma physics. As you know, plasma physics has started along two parallel lines. The first one was the hundred years old investigations in what was called electrical discharges in gases. This approach was to a high degree experimental and phenomenological, and only very slowly reached some degree of theoretical sophistication. Most theoretical physicists locked down on this field, which was complicated and awkward. The plasma exhibited striations and double-layers, the electron distribution was non-Maxwellian, there were all sorts of oscillations and instabilities. In short, it was a field which was not at all suited for mathematically elegant theories.

The other approach came from the highly developed kinetic theory of ordinary gases. It was thought that with a limited amount of work this field could be extended to include also ionized gases. The theories were mathematically elegant and when drawing the consequences of them it was found that it should be possible to produce a very hot plasma and confine it magnetically. This was the starting point of thermonuclear research.

However, these theories had initially very little contact with experimental plasma physics, and all the awkward and complicated phenomena which had been treated in the study of discharges in gases were simply neglected. The result of this was what has been called the thermonuclear crisis some 10 years ago. It taught us that plasma physics is a very difficult field, which can only be developed by a close cooperation between theory and experiments. As H. S. W. Massey once said (in a somewhat different context): «The human brain alone is not able to work out the details and understanding of the inner workings of natural processes. Without laboratory experiment there would be no physical science today.»

The cosmical plasma physics of today is far less advanced than the thermonuclear research physics. It is to some extent the playground of theoreticians who have never seen a plasma in a laboratory. Many of them still believe in formulae which we know from laboratory experiments to be wrong. The astrophysical correspondence to the thermonuclear crisis has not yet come.

I think it is evident now that in certain respects the first approach to the physics of cosmical plasmas has been a failure. It turns out that in several important cases this approach has not given even a first approximation to truth but led into dead-end streets from which we now have to turn back.

The reason for this is that several of the basic concepts on which the theories are founded, are not applicable to the condition prevailing in cosmos. They are «generally accepted» by most theoreticians, they are developed with the most sophisticated mathematical methods and it is only the plasma itself which does not «understand» how beautiful the theories are and absolutely refuses to obey them. It is now obvious that we have to start a second approach from widely different starting points.

4. Characteristics of first and second approach to cosmic plasma physics

The two different approaches can be summarized in Table 1.

If you ask where the border goes between the first approach and the second approach today, an approximate answer is that it is given by the reach of spacecrafts. This means that in every region where it is possible to explore the state of the plasma by magnetometers, electric field probes and particle analyzers, we find that in spite of all their elegance, the first approach theories have very little to do with reality. It seems that the change from the first approach to the second approach is the astrophysical correspondence to the thermonuclear crisis.

Table 1
Cosmical electrodynamics

First approach	*Second approach*
Homogeneous models	Space plasmas have often a complicated inhomogeneous structure
Conductivity $\sigma = \infty$ Electric field $E_{\parallel} = 0$	σ depends on current and often suddenly becomes 0, $E_{\parallel}$ often $\neq 0$
Magnetic field lines are «frozen in» and «move» with the plasma.	Frozen-in picture often completely misleading.
Electromagnetic conditions illustrated by magnetic field line picture.	It is equally important to draw the current lines and discuss the electric circuit.
Electrostatic double layers neglected.	Electrostatic double layers are of decisive importance in low density plasmas.
Filamentary structures and current sheets neglected or treated inadequately.	Currents produce filaments or flow in thin sheets.
Theories mathematically elegant and very well developed.	Theories still not very well developed and partly phenomenological.

5. *The origin of the solar system*

From what has been said it is obvious that astrophysics runs the risk of getting too speculative, unless it tries very hard to keep contact with laboratory physics. Indeed it is essential to stress that astrophysics is essentially an application to cosmic phenomena of the laws of nature found in the laboratory. From this follows that a particular field of astrophysics is not ripe for a scientific approach before experimental physics has reached a certain state of development. As a well-known historic example, before the advance of nuclear physics the attempts to understand how the stars generated their energy could not possibly be more than speculations without very much permanent value.

The problem of how the solar system originated has been the subject of a large number of highly divergent hypotheses. The reason for this has been that there was not enough basic knowledge of physics in the fields essential for the understanding of the phenomena and for a decision about which processes were possible.

However before we discuss any details of a theory of the origin and evolution of the solar system, it is essential to define what general character such a theory should have. In the past too much attention has been concentrated on the formation of planets around the sun. One of the unfortunate results of this is that many theories of the origin of the solar system have been based on theories of the early history of the sun. This is a very shaky basis because the formation of the sun (and other stars) is a highly controversial subject. Recognizing that the satellite systems of Jupiter, Saturn, and Uranus are very similar to the planetary system, and at least as regular as this system, it seems now more appropriate to aim at a general theory of the formation of secondary bodies around a central body, regarding the formation of the planetary system as only one of the applications of such a general theory.

The study of the sequence of processes by which the solar system originated has often been called *cosmogony*, a term which, however, is used in many other connections. As the origin of the solar system is essentially a question of the repeated formation of secondary bodies around a primary body, the term *hetegony* (from Greek *hetairos* or *hetes* = companion) has been suggested.

It seems likely (and is fairly generally agreed) that the sequence of events leading to the formation of the solar system is likely to have been as shown in Fig. 1 (we are here following what has been called the «planetesimal» approach). A primeval plasma was concentrated in certain regions around a central body, and condensed to small solid grains. (Even the primeval plasma

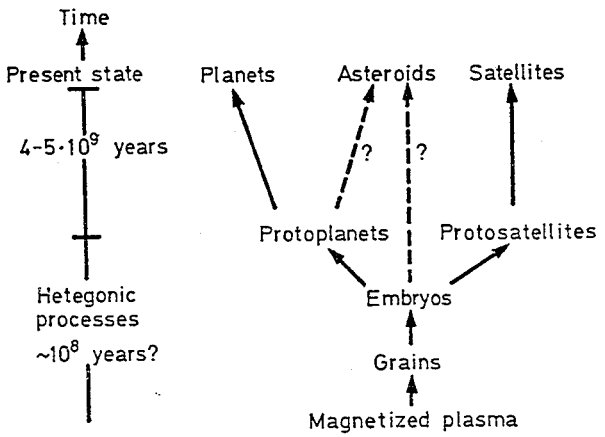


Fig. 1.

may have contained grains.) The grains accreted to what have been called embryos and by further accretion larger bodies were formed: planets if the central body was the sun, and satellites if it was a planet. The place of the asteroids in the hetegonic diagram is controversial. They have formerly been generally considered to be fragments of a broken-up planet, but there are now an increasing number of arguments for the view that they represent – or at least are similar to – an intermediate state in the formation of planets. A clarification of these two alternatives is important.

Even if the diagram of Fig. 1 is fairly generally accepted as it stands, this does not mean that the different processes are clarified. To a high degree they are still of a hypothetical character. Up to rather recently this has necessarily been the case because the basic processes have not been known very well. To some extent we have been in the same situation as the astrophysicists trying to clarify the energy generation in stars before the advent of nuclear physics. However, the situation seems now to be changing so that there is a good hope to bring the whole field of research from the state of a discussion of more or less bright hypotheses to a systematic scientific analysis.

6. Basic knowledge for the reconstruction of the hetegonic processes

Besides plasma physics, which we have already discussed, there are a number of other fields of research which are basic for the reconstruction of the hetegonic processes.

(1) *Plasma chemistry* means the field of research concerned with chemical reactions in a plasma. These are basically different from the reactions in non-ionized gases. It should also be considered to include the separation of different elements which takes place in an inhomogeneous plasma due to, *e.g.* temperature gradients and electric currents. Furthermore, the interaction between a plasma and a solid grain condensing from it is highly dependent on the state of ionization. The laboratory results and their application to cosmic conditions are relevant for the understanding of the different chemical composition of the celestial bodies.

For the next process in our evolutionary diagram, *viz.* the accretion of larger bodies from the initial condensation, the following fields of research are essential.

(2) *Solid-body collisions.* The grains which are the primary result of the condensation will move in Kepler orbits around the central body, but their motion will be disturbed by several effects. One of them is due to the mutual collisions. The relative velocities at these collisions may have any value from zero up to some 10 km/sec. This means that in many cases we are in the region of «hypervelocity» collisions. This is a field which is not yet understood very well. Available laboratory results seem to be scarce, and their application to cosmic conditions is uncertain because we know very little about the structure of the grains. Collisions between bodies with fluffy shock-absorbing surface layers are likely to differ from collisions between hard «marbles». Meteorite studies are supplying us with some information. The Apollo results about meteoroid impact on the lunar surface are another important source of knowledge. In these cases, however, we do not gain very much information about the structure of the grains in space, because the particles we recover have either passed the terrestrial atmosphere or been destroyed by impact on the lunar surface.

(3) The study *Kepler motion in a viscous medium* is essential for our understanding of the evolution of the orbits of the grains and the embryos. From a formal point of view this problem is similar to some basic problems in plasma physics, which are also concerned with a large number of interacting particles. It turns out that in the neighborhood of a central body the condensed grains have a tendency to move in similar orbits, thus forming what have been called «jet streams» in space.

(4) *Celestial mechanics* serves of course as a general background for the whole hetegonic process. This field has been rejuvenated by the application of computer analysis to many of the problems which were formerly impossible to

handle. Connected with this is the discovery of the importance of *resonance phenomena* in the present structure of the solar system. It seems likely that at hetegonic times also resonances played a decisive role.

(5) The hetegonic processes took place 4 to 5 billion years ago. The evolution of the primary product of these processes into our present-day solar system has consisted of a number of relatively slow changes: geological forces have transformed the structure of the planets, tidal effects have braked the spins of some of the bodies (especially of the satellites), collisions have taken place in the asteroidal belt, and there have been meteor impacts on the planetary surfaces, etc. All these effects are important for the reconstruction of the state of the system immediately after the hetegonic processes ended. It is only after «correcting» for them that the solar system data we observe today are of value for the reconstruction of the hetegonic processes.

7. Space observations relevant to the hetegonic problem

From the analysis we have made it is evident that the background knowledge necessary for the understanding of the hetegonic processes is rapidly increasing through advances in several different fields of research. We shall now discuss the question of what sort of space missions are of special value for the study of the hetegonic problem.

Let us first state that many of the space missions which are carried out today or planned for the future give valuable contributions. Increased knowledge of the behavior of cosmic plasmas is gained by spacecraft carrying out plasma and particle measurements in the magnetosphere and interplanetary space. Further, meteor impacts on spacecraft supply us with information of the very small bodies in our environment, which are probably related to those small bodies out of which our present planets once accreted. Particularly important is the study of meteor impacts on the Moon (and on Mars). Hence these and other investigations «automatically» contribute to the background knowledge necessary for the solution of the hetegonic problem. But although this is satisfactory there are a number of crucial problems which cannot be solved unless space research is purposely directed towards solving them. We shall now discuss how this could be done.

8. Big bodies versus small bodies

It is usually thought that after the lunar landings the most important missions will be those to Venus, Mars, and the other planets. This is not necessarily true, because missions to asteroids and comets would be at least as interesting from a scientific point of view. As some asteroids are the closest neighbours of the Earth-Moon system, this would also be the easiest from a technical point of view.

Our analysis has indicated which fields of research will contribute to the clarification of different phases of the hetegonic processes. Plasma physics and plasma chemistry are important for the first phase, including the condensation of small grains. The study of meteorite and asteroid sized bodies will have bearing on the accretion. We can state as a general rule that the smaller the body the further will the study of it bring us back in time. Thus small bodies will be relevant to earlier periods more than large bodies. This means that it is essentially through studying the properties of small bodies in space that we can hope to understand the crucial phase in the formation of the solar system when most of the matter, which later formed the planets and satellites, was still dispersed.

There is evidence that during the formation of the planets and satellites a great deal of information about the formation processes was stored in them. However, to a large extent this information is either obliterated or inaccessible. The planets are likely to have accreted from «planetesimals». The earliest phase of this accretion produced a small body, the matter of which may today be in the core of the planet, which means that it is inaccessible even if a manned spacecraft should land on the surface of the planet. There is also a possibility that, for example, convection in the interior of the planet has more or less completely obliterated the information once stored there. Concerning the surface layers, geological processes, including atmospheric effects, have mostly wiped out the surface traces of the hetegonic processes in Earth and probably also in Venus. In other bodies like the Moon and Mars, and probably also Mercury, there seems to be considerable information left, but only referring to the very last phase of the hetegonic processes.

Hence our conclusion is that studies of large bodies like the planets has only a limited value for the study of the origin of the solar system.

Asteroids, comets and meteoroids are different in this respect. Even if some of these bodies are fragments produced at collisions in space it is very likely that also these fragments contain considerable information about the conden-

sation and the accretion processes. Because of the smallness of the bodies there is no heating or convection in their interior which can obliterate the information stored from the time when they were produced, and at least in the very small bodies and by the fragmented bodies their «interior» is accessible. Furthermore, a study of them will give us knowledge of the behavior of small bodies in space which will be valuable for the clarification of the hetegonic processes in general. We study in them intermediate products in the manufacturing of planets. They give us, so-to-say, snapshots showing the sequence of events when a planet like the Earth once was created.

9. Old and new fields of science

We shall now return from our odyssey in both space and time to our starting point – how new technologies displace the center of gravity of the physical sciences. The great revolution in physics which took place in the beginning of this century meant that classical mechanics and classical electrodynamics were considered to be more or less obsolete as fields of research. The new fields which attracted the interest were the theory of relativity and quantum mechanics and the experimental work was largely concentrated on the exploration of the electron shells of the atom. The advance of nuclear physics marked another step in a similar direction.

The new trend which is introduced by the rise of plasma physics and space research is to some extent opposite. In these fields quantum mechanics and the theory of relativity are not very important. Instead, classical mechanics has become rejuvenated and is essential not only for calculating the trajectories of spacecraft but also for the study of the motion of the natural celestial bodies during their evolutionary history. Also classical electromagnetism is of decisive importance to the theory of magnetized plasmas, which is basic both for thermonuclear research and for astrophysics in general. This does not mean that we should make the mistake – similar to what was made 50 years ago – of declaring the atomic and nuclear physics to be obsolete. They are not. They have an enormous inertia which will keep them moving, and they will produce many new and interesting results. But they have got very serious competitors, and remarkably enough these are the fields which earlier were declared dead that are now being resurrected.

It is possible that this new era also means a partial return to more understandable physics. For the non-specialists four-dimensional relativity theory,

and the indeterminism of atom structure have always been mystic and difficult to understand. I believe that it is easier to explain the 33 instabilities in plasma physics or the resonance structure of the solar system. The increased emphasis on the new fields mean a certain demystification of physics. In the spiral or trochoidal motion which science makes during the centuries, its guiding center has returned to those regions from where it started. It was the wonders of the night sky, observed by Indians, Sumerians or Egyptians, that started science several thousand years ago. It was the question why the wanderers – the planets – moved as they did that triggered off the scientific avalanche several hundred years ago. The same objects are now again in the center of science – only the questions we ask are different. We now ask how to go there, and we also ask how these bodies once were formed. And if the night sky on which we observe them is at a high latitude, outside this lecture hall – perhaps over a small island in the archipelago of Stockholm – we may also see in the sky an aurora, which is a cosmic plasma, reminding us of the time when our world was born out of plasma. Because in the beginning was the plasma.

Biography

Hannes Olof Gösta Alfvén was born in Norrköping, Sweden, in 1908. His parents Johannes Alfvén and Anna-Clara Romanus were both practising physicians. Hannes Alfvén studied at Uppsala University from 1926, he obtained the degree of doctor of philosophy in 1934, in this same year he was appointed lecturer in physics at Uppsala University. In 1937 he became research physicist at the Nobel Institute for Physics in Stockholm, in 1940 he was appointed Professor in the Theory of Electricity at the Royal Institute of Technology in Stockholm, Professor of Electronics in 1945, and Professor of Plasma Physics in 1963. Since 1967 he is visiting professor of Physics at the University of California at San Diego.

In 1935 Hannes Alfvén married Maria Erikson Kerstin, they have five children: Cecilia, Inger, Gösta, Reidun and Berenike.

Professor Alfvén published a number of papers in physics and astrophysics, and the following monographs: Cosmical Electrodynamics, 1948; Origin of the Solar System, 1956; and together with C.-G. Fälthammar, Cosmical Electrodynamics, Fundamental Principles, 1963.

LOUIS NÉEL

Magnetism and the local molecular field

Nobel Lecture, December 11, 1970

1. The Weiss molecular field

It has long been known that ferromagnetism originates from interactions between the atomic magnetic moments which tend to align them parallel to one another in spite of thermal motion. In order to obtain a quantitative explanation of the experimental facts, Pierre Weiss[1] assumed that a ferromagnetic behaved as a pure paramagnetic, *i.e.* a paramagnetic with carriers of independent moments and magnetization given by

$$J = f(H/T) \tag{1}$$

and that the field of the interactions was equivalent to that of an imaginary magnetic field h_m, called the *molecular field*, which was proportional to the magnetization:

$$h_m = nJ \tag{2}$$

and which added to the applied field H. One thus obtains, if H is sufficiently the weak, so-called Curie–Weiss law of magnetization:

$$J = \frac{CH}{T-\theta} \text{ with } \theta = nC \tag{3}$$

If θ is positive, the susceptibility J/H becomes infinite when the temperature falls below the Curie point θ. From this temperature θ to absolute zero the substance assumes under the effect of its molecular field alone $h_m = nJ$ a certain *spontaneous magnetization*, J_s.

In this conception, the molecular field is considered to be a *uniform field* inside the ferromagnetic specimen. Furthermore, to give his theory a more complete nature, P. Weiss distinguished the *energy molecular field*, defined starting from the internal energy U by the relationship:

$$H_m = -(\partial U/\partial J) \tag{4}$$

from the *corrective molecular field of the equation of state* h_m defined, as stated

above, by:

$$J=f\left(\frac{H+h_m}{T}\right) \tag{5}$$

A thermodynamic argument shows that the following relationship exists between these two fields:

$$H_m = h_m - T\frac{\mathrm{d}h_m}{\mathrm{d}T} \tag{6}$$

This manner of explanation, theoretically very satisfactory, enables the energy properties of ferromagnetic substances to be treated and explained elegantly and simply, but it has the disadvantage of leading one to admit as a dogma the uniformity of the molecular field and all that follows therefrom, in particular the linear variation of the reciprocal susceptibility with temperature above the Curie point. This certainly retarded progress in the theory.

2. The local molecular field

On the other hand, P. Weiss was unable to give a satisfactory solution to the problem of the origin of the molecular field. It was only in 1928 that Heisenberg found an interaction mechanism giving a satisfactory order of magnitude. From the point of view of interest to us, it is essential only to point out that this involved very short-range interactions, preponderant between first-neighbour atoms and negligible beyond the second or third neighbours.

If, then, we consider an alloy composed of two kinds of randomly distributed atoms A and B, the surroundings of the atoms may be very different and the approximation of a unique molecular field representing for all sites the action of the surroundings must be very poor. The theoretical problem of the rigorous treatment of such interactions is still far from being solved but, while retaining the simplicity of the theories based on the molecular field, we can improve them considerably by introducing what I have called local molecular fields.

Weiss' hypothesis amounts to writing that the energy E_c of the system of A and B atoms is expressed in the form:

$$E_c = -\tfrac{1}{2}n(J_A+J_B)^2 \tag{7}$$

where J_A and J_B denote the magnetizations of the A and B atoms respectively. Actually, since their energy is the sum of the contributions made by pairs of

close-neighbour atoms, A–A, A–B, and B–B, we should rather write:

$$E_c = -\tfrac{1}{2}(n_{AA} J_A^2 + 2\, n_{AB} J_A J_B + n_{BB} J_B^2) \qquad (8)$$

This amounts to abandoning the notion of a general molecular field and to introducing *local molecular fields*, $h_A = n_{AA} J_A + n_{AB} J_B$, and $h_B = n_{AB} J_A + n_{BB} J_B$, acting on the A and B atoms respectively.

I developed this way of looking at the problem for the first time[2] in 1932, and showed that the susceptibility χ of an alloy containing proportions P and Q of A and B atoms, with Curie constants C_A and C_B, was expressed by:

$$\chi = \frac{T(PC_A + QC_B) - PQC_AC_B(n_{AA} + n_{BB} - 2\, n_{AB})}{T^2 - T(P\, n_{AA}\, C_A + Q\, n_{BB}\, C_B) + PQC_AC_B(n_{AA}\, n_{BB} - n_{AB}^2)} \qquad (9)$$

Instead of being represented by a straight line, the temperature dependence of the reciprocal susceptibility $1/\chi$ was now represented by a *hyperbola.*

I applied this theory to the interpretation of the properties of platinum–cobalt alloys and, a little later[3], to iron–cobalt, iron–nickel, and cobalt–nickel alloys.

At the time, this interpretation was not received with much favour. The existence of the straight Curie–Weiss lines in the $(1/\chi, T)$ representation was so well-established that when an experiment gave a curve it was rather preferred to break it down into a series of straight lines, each corresponding to a different magnetic state obeying the Curie–Weiss law.

3. The fluctuations of the Weiss molecular field

In Weiss' original theory the molecular field coefficient n and the Curie point θ both have positive values, and it is perfectly well understood that the molecular field, having a finite and positive value when the atomic moments are all parallel and in the same direction, can cause this ordered state. Weiss and his associates soon established that the paramagnetic properties of a number of salts were conveniently interpreted by a formula of type (3), but with a negative constant θ, *i.e.* with a negative molecular field. It could not be imagined how such a field could create an order at low temperatures.

On the other hand, Weiss' theory was incapable of explaining the properties of paramagnetic metals like manganese and chromium, which have a susceptibility almost independent of the temperature and too great to be attributable to a Pauli paramagnetism, *i.e.* due to electrons in an energy band.

Such was virtually the state of the problem in 1930, at which time I was

interested in the difference between the two Curie points, *i.e.* in the fact that the Curie point θ_p deduced from the Curie–Weiss law, or paramagnetic Curie point, differed from the ferromagnetic Curie point θ_f, the temperature at which the spontaneous magnetization disappeared, whilst Weiss' theory implies the equality $\theta_f = \theta_p$. To explain this difference I invoked the thermal fluctuations of the molecular field, the existence of which seemed certain, since this field arises from the action of the neighbour atoms. These thermal fluctuations are *time-varying fluctuations*, but in this way I also came very naturally to interest myself in *spatially varying fluctuations* and to analyse more closely the consequences of the elementary law of magnetic interaction, namely the existence of a coupling energy between two close-neighbour atoms, equal to $w \cos \alpha$, where α denotes the angle between their magnetic moments.

4. *Constant paramagnetism*

The constant w may be either negative or positive: negative in the case of ferromagnetism and positive in the case of a negative molecular field. In the latter case recourse to the molecular field, permissible at high temperatures when the locations of all atoms are on an average identical, is no longer so at low temperature, since the atomic magnets must then tend to group themselves by pairs of atoms with antiparallel moments.

In this way I noted (ref. 5, p. 64) that in a *bcc* lattice, composed of two interlaced simple cubic sub-lattices, stable equilibrium at low temperature corresponded to an orientation of the atomic moments in one of the sub-lattices in a certain direction and an orientation in the other direction of the atomic moments in the other sub-lattice, as shown in Fig. 1 for the case of a plane lattice. This assembly deforms (Fig. 2) under the action of a magnetic field, and assumes an average induced magnetization, given for one atom by

$$\bar{\mu} = \frac{\mu^2 H}{6pw} \tag{10}$$

where μ denotes the atomic magnetic moment and $2p$ is the number of neighbours. In this way a *constant paramagnetism* is obtained, *i.e.* a susceptibility independent of the field and of the temperature.

However, at high temperatures, we should observe a Curie–Weiss law of type (3).

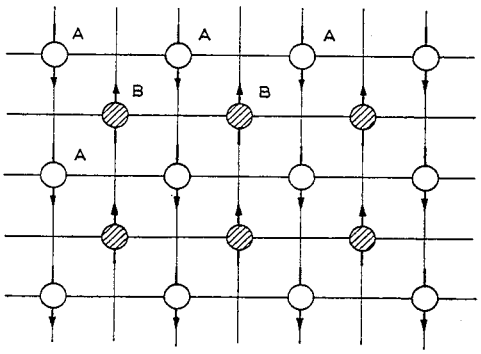


Fig. 1. Resolution of a plane lattice into two sub-lattices magnetized in antiparallel.

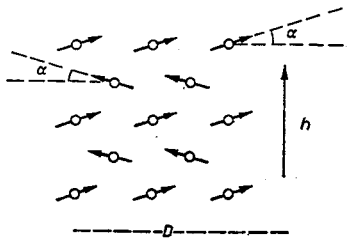


Fig. 2. Deformation of the antiparallel arrangement of atomic magnetic moments under the action of a field h.

In order to study the transition from constant paramagnetism to Curie–Weiss paramagnetism, I assumed for the sake of simplicity that the atoms could be grouped by pairs independent of one another, the internal energy of a pair being taken as $w \cos \alpha$. This led to the variation

$$\bar{\mu} = \frac{\mu^2 H}{3w}\left[1 - \frac{2w}{kT}\,\frac{1}{e^{2w/kT}-1}\right] \qquad (11)$$

The temperature dependence of the reciprocal susceptibility is shown by the curve in Fig. 3. As T is raised from 0 to θ, the susceptibility varies by only 1.4%.

In another paper, which appeared almost at the same time[4], I proposed interpreting the constant paramagnetism of manganese and chromium by the previous mechanism, with values of θ respectively equal to 1720 °K and 4150 °K. To support this interpretation, I showed that on diluting manganese and chromium by copper, silver or gold, a progressive change from a constant paramagnetism to a Curie–Weiss paramagnetism was observed, as

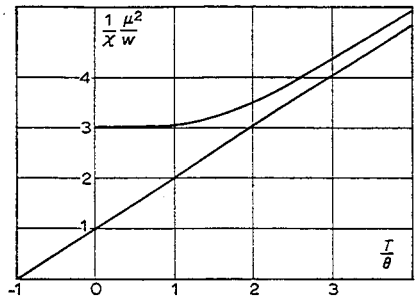


Fig. 3. Temperature dependence of the reciprocal susceptibility of an assembly of randomly-oriented pairs of atoms with anti-parallel moments.

might have been expected from the previous theory when the interactions, *i.e.* θ, are decreased.

In 1936, I again took up the theory of negative interactions[5], in the low-temperature region, allowing for the *coupling of the atomic moments with the crystal lattice*. The coupling energy was taken as $w'' \cos 2\theta$, where θ denotes the angle between the atomic magnetic moment and a privileged direction D of the lattice. It was then shown that the magnetic susceptibility is a complex function of the magnetic field H and the angle β that H makes with D, shown in Fig. 4 for various values of β at 10-degree intervals.

In particular, if the field is parallel to the privileged direction, the susceptibility, initially zero in weak fields, suffers a discontinuity for

$$H = H_0 = (8pww''/\mu^2)^{\frac{1}{2}}$$

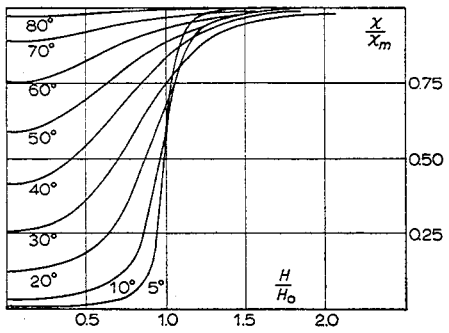


Fig. 4. Variation of the susceptibility of an assembly of atoms with antiparallel moments with the magnetic field and the direction of alignment of the magnetic moments with respect to the crystal lattice.

and takes a value $\chi_m = (\mu^2/4pw)$ independent of H: when H reaches the value H_0, the moments which were originally parallel to D and to H orient themselves abruptly in a perpendicular direction. Fifteen years later, in 1951, C.J. Gorter and his associates[5] observed this phenomenon for the first time in copper chloride, $CuCl_2 \cdot 2H_2O$, at 4.1 °K.

It was also deduced from this theory that in a polycrystalline sample, in which the directions D are randomly distributed, the susceptibility is weaker in weak fields than in strong fields, and varies schematically with temperature as shown in Fig. 5. It will be noted that there is no transition temperature: the change from constant paramagnetism to Curie–Weiss paramagnetism takes place perfectly smoothly.

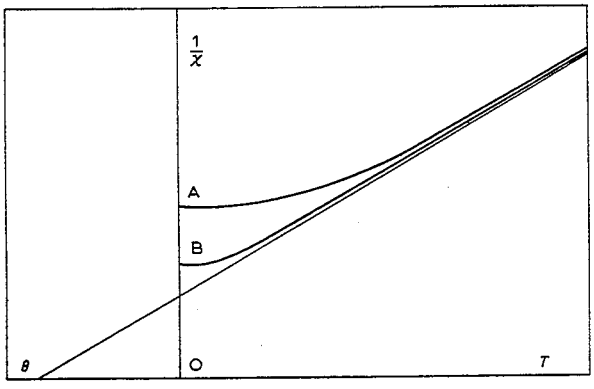


Fig. 5. Influence of the magnitude of the magnetic field on the susceptibility of an assembly of atoms with anti-parallel moments.

5. *Discovery of the antiferromagnetic transition point*

Still in 1936, I had the idea[5] of applying the theory of the local molecular field to the two sub-lattices A and B with magnetizations J_A and J_B used in the previous papers, and of representing the interactions by imaginary fields H_A and H_B with, at low temperature and for $H = 0$, the fundamental relationship $J_B = -J_A$.

The result was that the two sub-lattices had to acquire spontaneous magnetizations in opposite directions, disappearing at a certain transition temperature θ_N, known nowadays as the Néel temperature, following a proposal made by C.J. Gorter. We were hence faced with a new kind of magnetic material,

composed of the sum of two interlaced identical ferromagnetics spontaneously magnetized in opposite directions. Effects depending on the square of the spontaneous magnetization, such as the specific heat anomaly, should thus show the same variation as in ferromagnetic substances.

In the absence of coupling with the crystal lattice, the susceptibility remains constant in the interval $0 < T < \theta_N$ and then obeys a Curie–Weiss law for $T > \theta_N$, without suffering a discontinuity at the Curie point.

Two years later, Squire, Bizette and Tsaï[8] discovered that MnO possessed the predicted properties, in particular a transition temperature at $\theta_N = 116$ °K. Much later, in 1949, C. G. Shull and S. J. Smart[9] using a neutron diffraction, confirmed that the atomic moments effectively possessed the antiparallel orientations predicted by the sub-lattice theory.

F. Bitter then completed the theory[10] by calculating as far as the transition point θ_N on the susceptibility in a magnetic field parallel to the direction Δ of alignment of the antiparallel moments, and gave the name of antiferromagnetics to this new category of magnetic substances; in 1941 J. H. van Vleck[11] collected and reviewed the results obtained.

Finally, the situation may be summed up as follows, according to the orientation of the magnetic field H with respect to the antiferromagnetism direction Δ, *i.e.* the privileged direction of the crystal lattice to which the atomic mag-

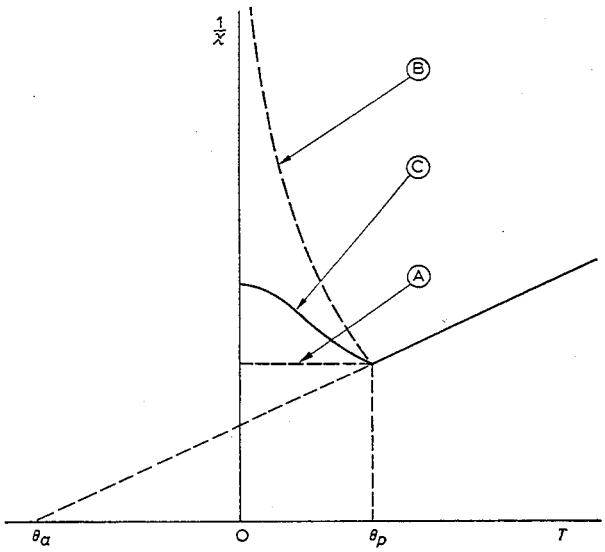


Fig. 6. Temperature dependence of the reciprocal susceptibility of an antiferromagnetic material.

netic moments are parallel, in one direction or the other, in the absence of an external applied field. When H is perpendicular to Δ, one obtains[7] curve A in Fig. 6. When H is parallel to Δ, curve B is obtained[10]. Finally, for a polycrystalline substance, the intermediate curve C is obtained, demonstrating at the Néel point θ_p the existence of a sharp maximum in the susceptibility, marked also by a sharp maximum in the specific heat. At temperatures below this maximum an antiferromagnetic order exists, which may be detected by neutron diffraction.

A large number of antiferromagnetic materials is now known: these are generally compounds of the transition metals containing oxygen or sulphur. They are extremely interesting from the theoretical viewpoint, but do not seem to have any applications.

6. Other investigations in the field of magnetism

It should be pointed out that the theoretical studies which I had developed from 1930 to 1938 had been carried out with the aim of determining the elementary interaction energies w between two close-neighbour magnetic atoms, whether similar or not. I expected to find a universal law giving w as a function of the distance between the interacting atoms alone, but this was obviously a rather oversimplified view and my expectations came to nothing. Nevertheless, positive results had been obtained and had made it possible to find interesting correlations between the anomalies of expansion, the effect of pressure, and the variation of the molecular field with temperature. Even now these questions are still on the agenda.

At all events, in 1938 I abandoned the problem of interactions and the local molecular field for about eight years to devote myself to other subjects such as the Rayleigh laws, fine particles, and the role of internal dispersion fields, to quote only those which led to the most important results.

The protection of ships against magnetic mines by a method I proposed, consisting of giving them a permanent magnetization equal and opposite to the magnetization induced by the terrestrial magnetic field, drew my attention to the laws of magnetization of ferromagnetics in weak fields, known as the Rayleigh laws. I was able to give an interpretation of these[12] based on the propagation of the Bloch walls in a randomly perturbed medium.

I also systematically studied the part played by the internal demagnetizing or dispersion fields, up to that time almost neglected. It was then possible to

deduce a more correct theory of the laws of approach of technical magnetization to saturation taking into account the random orientation of crystallites[13] and the presence of cavities or non-magnetic inclusions[14], and to give a more general theory of the coercive force taking into account the irregularities of internal mechanical tensile stresses as well as the presence of cavities or non-magnetic inclusions[15]. On the same general lines, I developed a theory of magnetization for ferromagnetic single crystals[16] along successive modes, each specified by the number of *phases, i.e.* by the number of different categories of elementary domains with parallel spontaneous magnetizations. This theory, in complete agreement with the experimental facts, put the final full-stop to a subject which up to then had been very poorly understood.

Finally, I showed as far back as 1942 that sufficiently small ferromagnetic particles must enclose only a single elementary domain and, depending in their dimensions, must behave either as a superparamagnetic substance or as a set of small permanent magnets and show macroscopically a hysteresis loop with a high coercive force. In this way good permanent magnets can be made from particles of soft iron having only a shape anisotropy. These properties and their applications were developed by L. Weil. The publication of all these results was delayed until 1947[17] because of the German occupation and for reasons of patent rights.

7. *Magnetic properties of the spinel ferrites*

It was in 1947, after reading a paper by Verwey and Heilmann[18] on the structure of ferrites, that I returned to the study of these substances. These ferrites Fe_2O_3MO, with a spinel structure, where M is a divalent metal, are divided into two categories according to their magnetic properties: *paramagnetic ferrites*, in which M represents zinc or cadmium, and *ferromagnetic ferrites* in which M represents manganese, cobalt, nickel, etc. The latter have a considerable technical interest, for these ferromagnetics are electrical insulators, but their theoretical interest is also high in view of their curious properties, very different from those of the classical ferromagnetics: their saturation molecular moment of 1 to 5 μ_B (μ_B is the Bohr magneton), is much weaker than the total magnetic moment of the ions contained in the molecule, which varies from 10 to 15 μ_B. In addition, above the Curie point, the temperature dependence of the reciprocal susceptibility has a quite extraordinary hyperbolic shape, concave downwards, towards the temperature axis, and with a high

temperature asymptote extrapolating back towards a negative absolute temperature.

Mlle Serres, to whom we owe some fine experimental work on ferrites, interpreted[19] the shape of the $(1/\chi, T)$ curve by the superposition of a temperature-independent paramagnetism due to the ferric ions, equal to that of α-Fe_2O_3. But as this paramagnetism is not an atomic property, we cannot see why the ferric ions should retain in the ferrites the same constant paramagnetism as in α-Fe_2O_3.

With regard to the crystalline structure, the metal ions occupy, in the interstices of a close-packed cubic lattice of oxygen ions, sites A surrounded by four O^{2-} ions and sites B surrounded by six O^{2-} ions. There are two categories of ferrites: *normal ferrites*, in which the two Fe^{3+} ions of the molecule occupy the two sites B and the M^{2+} ion site A, and *inverse ferrites*, where one of the Fe^{3+} ions occupies site A and the other one of the two sites B. From an X-ray examination, Verwey and Heilmann concluded that the ferromagnetic ferrites are inverse and paramagnetic ferrites are normal.

8. Foundation of the theory of ferrimagnetism

To interpret the magnetic properties, I assumed[20] that the predominant magnetic interactions were exerted between the ions placed at sites A and ions placed at sites B, and that they were *essentially negative*. At absolute zero, these strong negative interactions make the magnetic moments of the A ions align themselves parallel to one another to give a resultant moment $\mathbf{M}_{as}$ pointing in the opposite direction to the resultant $\mathbf{M}_{bs}$ of the magnetic moments of the B ions, these also being all parallel. The observable spontaneous magnetization is equal to the difference $|\mathbf{M}_{as}-\mathbf{M}_{bs}|$.

To study the consequences of these hypotheses, I applied to the two sub-lattices A and B the concept of the local molecular field developed earlier[2]. Letting λ and μ stand for the proportions of magnetic ions, all assumed to be identical ($\lambda+\mu=1$), located in the A and B sub-lattices respectively, and designating by $-n$, $n\alpha$ and $n\beta$ the molecular field coefficients corresponding to the interactions AB, AA, and BB respectively, the discussion showed that the variation of the spontaneous magnetization with temperature would have to assume the rather extraordinary forms shown diagrammatically in Fig. 7, in which the capital letters correspond to the regions of the (α, β) plane shown in Fig. 8: region G at the bottom left corresponds to paramagnetism. All

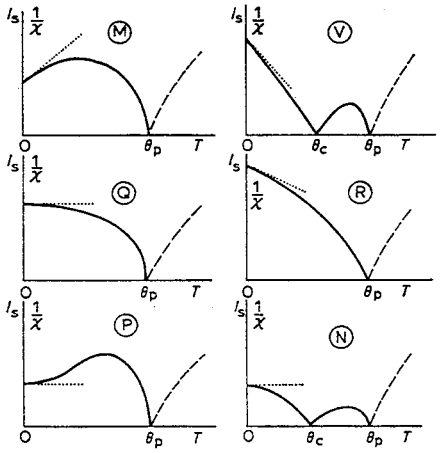


Fig. 7. Temperature dependence of the spontaneous magnetization and the reciprocal susceptibility for a ferrimagnetic material: the various types possible.

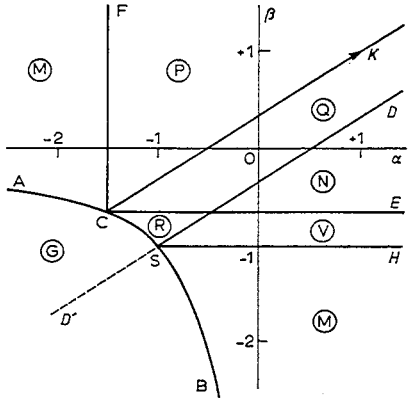


Fig. 8. Possible curve types in Fig. 7, according to the values of α and β.

the curve types illustrated in Fig. 7 were subsequently found experimentally.

Above the Curie point θ_p the ferrite becomes paramagnetic, with a susceptibility χ such that:

$$\frac{1}{\chi} = \frac{T}{C} + \frac{1}{\chi_0} - \frac{\sigma}{T-\theta} \qquad (12)$$

with the notations:

$$\frac{1}{\chi_0} = n\,(2\,\lambda\mu - \lambda^2\alpha - \mu^2\beta)$$

$$\sigma = n\,C^2\,\lambda\,\mu[\lambda(1+\alpha) - \mu(1+\beta)]^2$$

$$\theta = n\,C\,\lambda\,\mu(2+\alpha+\beta)$$

We are hence confronted by substances whose ferromagnetism is due to *negative interactions*, which is quite a remarkable fact, and whose properties, appreciably different from those of classical ferromagnetics, justify a special name: I proposed calling them *ferrimagnetics*.

9. *Comparison of the theoretical and experimental results*

This theory immediately makes it possible to interpret[20] the properties of magnesium, lead, and calcium ferrites, those of magnetite Fe_3O_4, and of manganese antimonide Mn_2Sb. As examples, Figs. 9–12 show a comparison of the experimental and the calculated values in the case of Fe_3O_4 and of Mn_2Sb for the temperature dependence of the spontaneous magnetization and of the reciprocal susceptibility. Naturally, the same set of coefficients n, α and β is used in both cases. The result was encouraging, so a considerable experimental effort was launched immediately to test the theory, since the existing experimental material was fairly meagre.

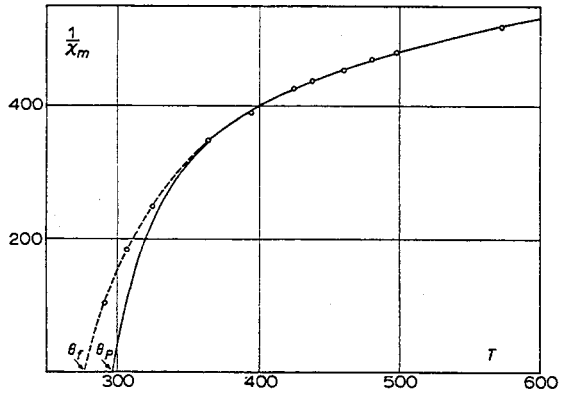


Fig. 9. Temperature dependence of the reciprocal susceptibility of magnetite (fitted curve and experimental points). Note the difference between the ferromagnetic and paramagnetic Curie points.

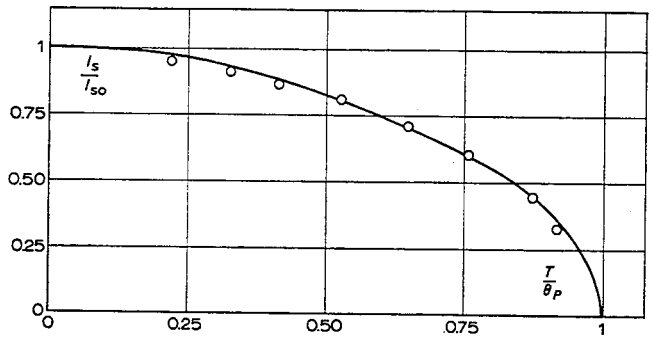


Fig. 10. Temperature dependence of the spontaneous magnetization of magnetite (experimental points and curve calculated with the molecular-field coefficients deduced from the study of susceptibility).

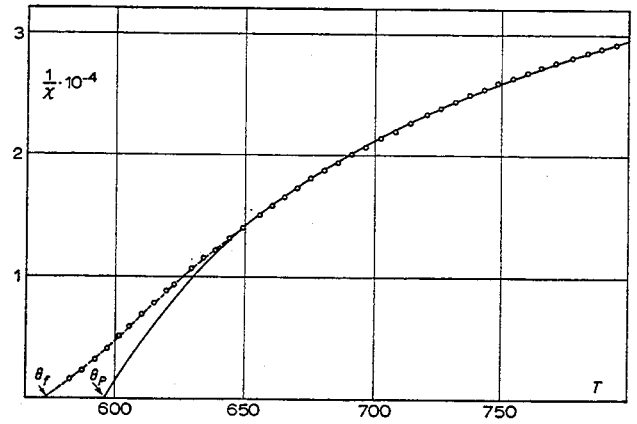


Fig. 11. Temperature dependence of the susceptibility of manganese antimonide (experimental points and fitted curve).

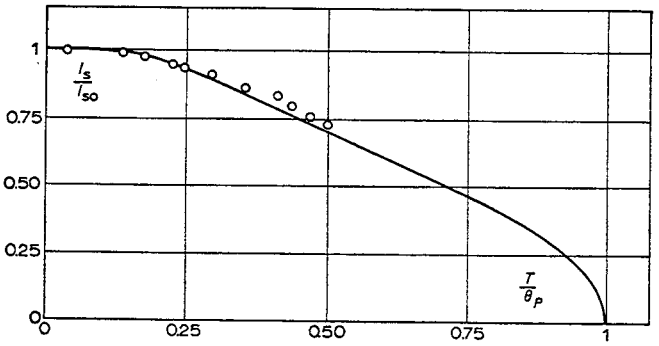


Fig. 12. Temperature dependence of the spontaneous magnetization of manganese antimonide Mn_2Sb (experimental points and curve calculated with the molecular field coefficients deduced from the study of susceptibility).

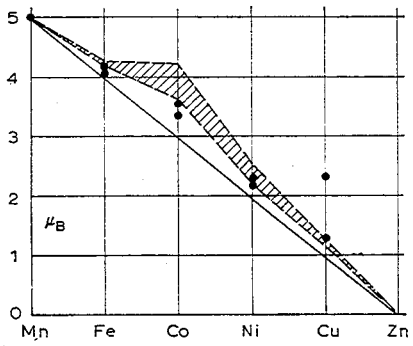


Fig. 13. Saturation magnetization of certain ferrites at absolute zero, for various M ions.

In *pure inverse ferrites* Fe_2O_3MO, with M = Mn, Fe, Co, Ni, the saturation magnetic moment at absolute zero must simply be that of the M ion, since the magnetic moments of the two Fe^{3+} ions exactly cancel each other, located as they are on different kinds of sites. Fig. 13 enables the theory to be compared with experiment. The points correspond to the experimental results[21]. The straight line corresponds to the theoretical predictions with the «spin only» values of the magnetic moments, and the shaded region to the theoretical predictions corrected by taking into account the incompletely quenched orbital angular moment, deduced from the determination of the effective moment of the corresponding paramagnetic salts. The agreement is very satisfactory. With regard to the copper ferrite, experiment shows that the saturation magnetic moment depends on its heat treatment. Cooled slowly, this ferrite assumes the inverse structure whilst at high temperatures the Fe^{3+} ions are distributed at random over the sites A and B, for the difference in energy between the normal structure and the inverse structure is slight, of the order of magnitude of kT. A detailed analysis of the phenomenon shows that this interpretation is correct and hence lends support to the theory of ferrimagnetism[22].

For the same ferrites, it has also been possible to interpret appropriately the experimental results on the temperature dependence of the spontaneous magnetization[23] and of the susceptibility above the Curie point[24]. Figs. 14 and 15, relating to the three ferrites of iron, cobalt, and nickel, show the agreement between the theory and experiment.

Over and above this, on studying magnetite by neutron diffraction, Shull *et al.*[25] confirmed that the magnetic moments of the atoms placed at sites A were effectively pointed in the direction opposite to the moments of the atoms at sites B.

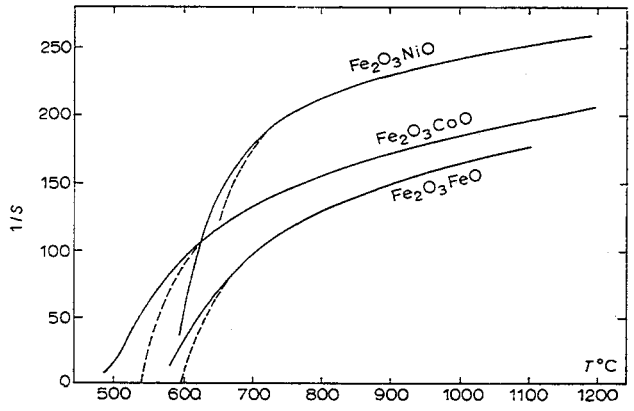


Fig. 14. Temperature dependence of the inverse susceptibility of different ferrites (experimental curves shown in full lines, calculated curves dotted).

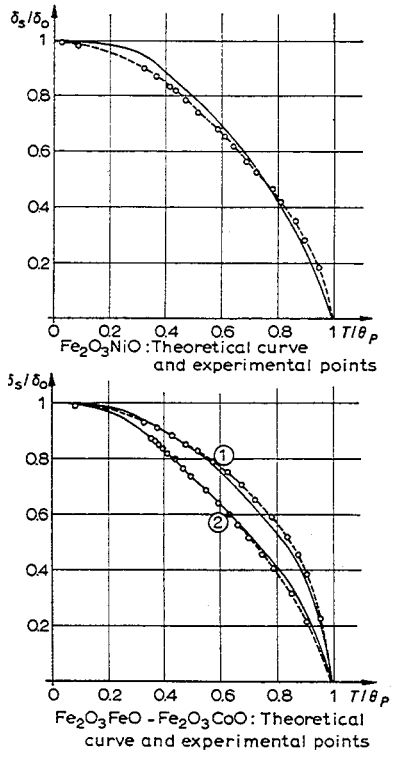


Fig. 15. Temperature dependence of the spontaneous magnetization of various ferrites (experimental points and curves calculated with the molecular-field coefficients deduced from the study of the susceptibility).

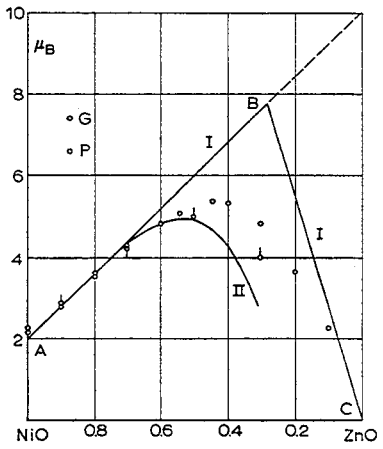


Fig. 16. Variations of the saturation magnetization of composite nickel and zinc ferrites with the zinc content (experimental points and calculated curves).

The theory of ferrimagnetism also makes it possible to explain the behaviour of composite ferrites, like the nickel–zinc ferrites with the formula $Fe_2N_{1-x}Zn_xO_4$. An atom of zinc which replaces an atom of nickel will locate itself on an A site, since that is the place preferred by zinc, simultaneously, the Fe^{3+} ion located on this A site will occupy at B the place left vacant by the departure of the Ni ion at the same time as its magnetic moment of $5\,\mu_B$ reverses direction. The net variation of the saturation magnetic moment is therefore equal to $8\,\mu_B$, the difference between $10\,\mu_B$ resulting from the reversal and $2\,\mu_B$ resulting from the departure of the Ni^{2+} ion. The slope of the initial tangent to the curve giving the saturation magnetization as a function of x is therefore equal to $8\,\mu_B$. The experimental results confirm this prediction[26] (Fig. 16).

10. The case of pyrrhotite

Shortly afterwards, the same theory made it possible to solve the riddle posed by pyrrhotite, Fe_7S_8, a ferromagnetic compound studied a long time ago by P. Weiss[27], which has a small saturation magnetic moment, of the order of 3 μ_B, whilst from its formula a value of about $30\,\mu_B$ might be expected. Pyrrhotite has the same crystal structure, Ewald's type B8, as the sulphide FeS, which is a typical antiferromagnetic: in the latter the successive planes of iron atoms, perpendicular to the ternary axis, are magnetized in a certain direction in one

sense and the other in alternation. F. Bertaut has shown[28] that actually pyrrhotite is a compound containing vacancies whose formula should rather be written Fe_7S_8T, where T denotes vacancy, *i.e.* a site that would be occupied by an Fe^{2+} ion in the compound FeS, but that is vacant in Fe_7S_8. At low temperature the vacancies assume an ordered distribution and group themselves, as shown in Fig. 17, on the even-order iron planes: in short, the vacancies tend

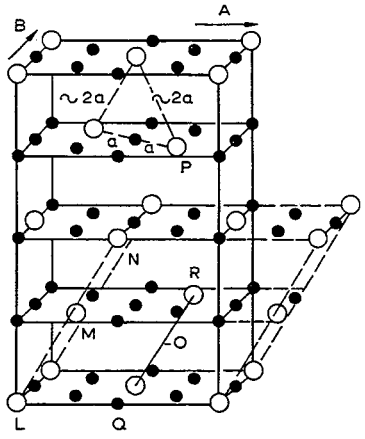


Fig. 17. Crystal structure of pyrrhotite at low temperature. The black circles represent the iron atoms and the white circles the vacancies. The odd-order iron planes and the oxygen atoms are not shown.

to space themselves as far as possible from one another. On the odd-order iron planes all the sites are occupied. The two sub-lattices equivalent in the case of FeS become different, and we are then dealing in the case of Fe_7S_8 with true ferrimagnetism. When the temperature rises above 900 °K, the ordering of the vacancies disappears and the two sub-lattices become statistically equivalent: correlatively, the paramagnetic susceptibility of pyrrhotite assumes values close to those of FeS.

11. Discovery of the garnet ferrites

There is no doubt that the garnet ferrites offer the finest illustration of ferrimagnetism. These substances, with the general formula $Fe_5M_3O_{12}$, where M is a trivalent rare-earth metal ion, form an important class of magnetic compound whose properties are explained simply and accurately by ferrimagne-

tism with three sub-lattices. Besides, since the metal ions entering into their composition can be replaced by a very wide variety of other ions, these compounds are of great interest in the theoretical study of interactions. From the point of view of applications, they are excellent insulators, can be prepared in large crystals, and have very sharp resonance lines; they can be used at very high frequencies in a large number of devices.

The history of their discovery begins in Strasbourg in 1950, where Forestier and Guiot-Guillain[29] heated an equimolecular mixture of Fe_2O_3 and M_2O_3 (M = rare earth metal) and obtained strongly ferromagnetic products whose Curie points ranged from 520 to 740 °K and to which they attributed a perovskite-type structure[30]. They then demonstrated the curious fact[31] that with M = Yb, Tm, Y, Gd, or Sm, these products had two Curie points spaced about a hundred degrees apart and varying regularly as a function of the atomic radius of M.

These results attracted the attention of research workers at Grenoble: Pauthenet and Blum[32] prepared a gadolinium ferrite and showed that besides the two Curie points, $\theta_1 = 570$ °K and $\theta_2 = 678$ °K, this compound had a third transition temperature $\theta_3 = 306$ °K, identifiable as a *compensation temperature* in the sense of the ferrimagnetism theory, *i.e.* as a temperature corresponding to a change of sign of the spontaneous magnetization (types V and N in Fig. 7).

To explain these facts, I suggested[33] that the Fe^{3+} ions had to form a ferrimagnetic arrangement A, with a structure independent of M, whose resultant spontaneous magnetization magnetized the B sub-lattice of the M^{3+} ions in the opposite direction. The molecular field h_A representing the action of A on B had to be sufficiently weak for the magnetization of B to be effectively equal to $C_B h_A/T$, at least above 100 °K, the temperature θ_3 being one at which the magnetization of B is equal and opposite to that of A. In particular, it followed from this that the temperature θ_3 had to be the lower the smaller is the Curie constant C_B of the M^{3+} ions. In agreement with these predictions, it was shown a few days later[34] that the ferrites of dysprosium and erbium also had compensation temperatures, located at 246 and 70 °K respectively. Despite these successes, this assumed structure was incompatible with that of a perovskite with formula $FeMO_3$.

Starting from the assumption of another compound mixed with the perovskite, Bertaut and Forrat[35] showed in January 1956 that it was a question of a cubic compound, $Fe_5M_3O_{12}$, of space group O_h^{10} 1a3d, with 8 molecules in the unit cell, of structure identical to that of the semi-precious stones known as *garnets* (Fig. 18). The primary ferrimagnetic arrangement comprises 24 Fe^{3+}

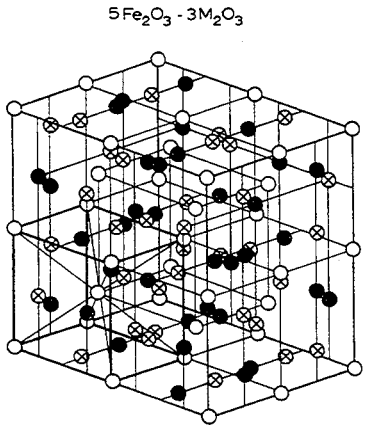


Fig. 18. Crystal structure of the garnet ferrites. ○, Fe ion on site a; ⊗, Fe ion on site d; ●, M ion on site c. (after Strukturbericht and F. Forrat)

ions on sites d surrounded by 4 O^{2-} ions, and 16 Fe^{3+} ions on sites a, surrounded by 6 O^{2-} ions. This arrangement bound by strong interactions, has magnetic properties effectively independent of the nature of the ions M, to which it is coupled antiferromagnetically be weak interactions. It can be studied magnetically in the isolated state by taking a non-magnetic ion such as Y or Lu for M.

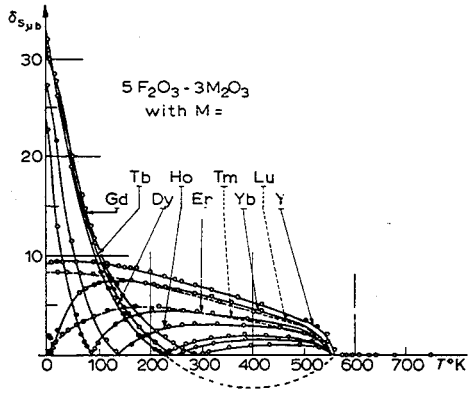


Fig. 19. Temperature dependence of the spontaneous magnetization of a series of garnet ferrites. In particular, it will be noted how the compensation point varies with the atomic number of the rare earth.

12. Interpretation by ferrimagnetism with three sub-lattices

The generalization of the theory of ferrimagnetism to 3 sub-lattices and the representation of the interactions by means of 9 local molecular fields, 6 of them independent, raises no problem. The agreement of the calculated curves with the experimental results is very satisfactory, as shown in Fig. 19, taken from Pauthenet's fundamental paper[36] on the temperature dependence of the spontaneous magnetization of various garnets. The near identity of their Curie points well illustrates the fact that the nature of the M ions does not modify the ferrimagnetic arrangement of the Fe^{3+} ions. Above the Curie point, in the paramagnetic region, the agreement of the calculated curves with the experimental curves is also satisfactory (Fig. 20).

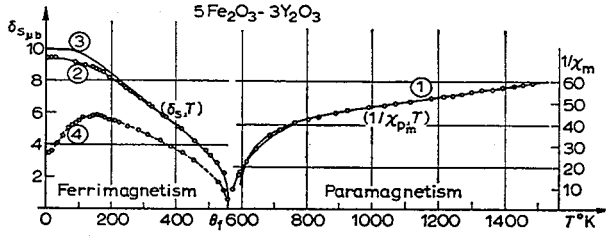


Fig. 20. Yttrium garnet. Temperature dependence of the spontaneous magnetization and of the inverse susceptibility; experimental points and curves calculated using the same molecular field coefficients. Curve (4) refers to an yttrium garnet containing a small amount of gadolinium.

13. Superexchange

Both in respect of antiferro- and ferrimagnetism, the theory makes it possible to deduce from the experimental curves the elementary interaction energy w between two close-neighbour magnetic atoms. This calculation undertaken for spinel ferrites and a few antiferromagnetics such as MnO or NiO brings out some new facts[20,37]. In these substances the interactions between neighbouring iron ions are negative, at exactly the mutual distances for which they are positive in the pure metal or in alloys. It is further observed that there are considerable interactions between magnetic atoms separated by oxygen atoms, which ought to act as screens. The answer is even more revolutionary: it is actually a case of *superexchange interactions*, in which the oxygen atom separating two iron atoms plays an essential part. A considerable time ago

Kramers had already predicted[38] the possibility of such interactions, but this was the first time that their objective existence had been clearly demonstrated.

Thus, in oxides and ferrites exchange interactions of the classical type M–M coexist with superexchange interactions of the type M–O–M; this, then, is first-class material for testing the theories of the various interactions.

14. Conclusions

Despite its rather naive simplicity, the local molecular-field method has had some undeniable successes in linking up, in an intellectually very satisfying way, a large number of already known facts and in leading to the discovery of new facts.

It should be noted that all the structures discussed in this paper are *collinear*: all the atomic magnets are on average (in time) parallel, in one sense or other, to the same direction. However, the method of the local molecular field may be extended also to non-collinear structures, such as that of *helimagnetism*, which Yoshimori and Villain discovered independently and absolutely unexpectedly: in this way the phenomena can be interpreted remarkably simply and specifically.

In spite of everything, it hardly seems recommendable to extend the method to more complex structures, such as the umbrella structure, requiring the main crystal lattice to be broken down into a large number of sub-lattices. Indeed, under these conditions, an atom belonging to a given sub-lattice has only a very small number of neighbours in each of the other sub-lattices, often it has only one or two. The molecular field method, which consists of replacing the instantaneous action of an atom by that of an *average atom*, has many more chances of leading to a correct result as it deals with a greater number of atoms. It is probably the more correct, the higher the atomic spin number. Independently of this problem, the method applied to a large number of sub-lattices loses its main advantage of simplicity.

The method also has more insidious traps. If the parameters are suitably chosen, it can lead to the calculation of curves showing the temperature dependence of the spontaneous magnetization, or of the paramagnetic susceptibility, which coincide remarkably with the experimental results, say to within a few parts in a thousand. Under these conditions it could be expected that the elementary interaction energies deduced from these parameters would correspond to the true values with the same accuracy. Nothing is farther from

the truth: errors of 10–20% and even more are made. Some prudence is therefore indicated.

On the other hand, recourse to the local molecular field seems essential, for the most rigorous methods lead to inextricable complications. It should be remembered that we still do not have the rigorous solution to even the simplest case, this being the case of a single cubic lattice with identical atoms of spin 1/2 and interactions reduced to those exerted between close-neighbour atoms. What, then, are we to think of the case of garnets with 160 atoms in the unit cell, spins of up to 5/2, and at least six different coupling constants? The imperfections in molecular field methods must be treated with indulgence when we consider the simplicity with which the successes discussed in the first lines of these conclusions were obtained.

1. P. Weiss, *Compt. Rend.*, 143 (1906) 1137; *J. Phys. (Paris)*, 6 (1907) 666.
2. L. Néel, *Ann. Phys. (Paris)*, 17 (1932) 5.
3. L. Néel, *Compt. Rend.*, 198 (1934) 1311.
4. L. Néel, *J. Phys. (Paris)*, 3 (1932) 160.
5. L. Néel, *Ann. Phys. (Paris)*, 5 (1936) 232.
6. N. J. Poulis, J. van den Haendel, J. Ubbink, J. A. Paulis and C. J. Gorter, *Phys. Rev.*, 82 (1951) 52.
7. L. Néel, *Compt. Rend.*, 203 (1936) 304.
8. C. Squire, H. Bizette and B. Tsaï, *Compt. Rend.*, 207 (1938) 449.
9. C. G. Shull and S. J. Smart, *Phys. Rev.*, 76 (1949) 1256.
10. F. Bitter, *Phys. Rev.*, 54 (1938) 79.
11. J. H. van Vleck, *J. Chem. Phys.*, 9 (1941) 85.
12. L. Néel, *Cahiers Phys.*, No. 12 (1942) 1; No. 13 (1943) 18.
13. L. Néel, *Compt. Rend.*, 220 (1945) 814.
14. L. Néel, *Compt. Rend.*, 220 (1945) 738.
15. L. Néel, *Cahiers Phys.*, No. 25 (1944) 21; *Compt. Rend.*, 223 (1946) 198; *Ann. Univ. Grenoble*, 22 (1946) 299.
16. L. Néel, *J. Phys. (Paris)*, 5 (1944) 241, 265.
17. L. Néel, *Compt. Rend.*, 224 (1947) 1488; 224 (1947) 1550; 225 (1947) 109.
18. E. J. W. Verwey and E. L. Heilmann, *J. Chem. Phys.*, 15 (1947) 174.
19. A. Serres, *Ann. Phys. (Paris)*, 17 (1932) 53.
20. L. Néel, *Ann. Phys. (Paris)*, 3 (1948) 137.
21. P. Weiss and R. Forrer, *Ann. Phys. (Paris)*, 12 (1929) 279; E. W. Gorter, *Compt. Rend.*, 230 (1950) 192; C. Guillard, *Compt. Rend.*, 229 (1949) 1133.
22. L. Néel, *Compt. Rend.*, 230 (1950) 190.
23. R. Pauthenet and L. Bochirol, *J. Phys. (Paris)*, 12 (1951) 249.
24. M. Fallot and P. Maroni, *J. Phys. (Paris)*, 12 (1951) 256.

25. C. G. Shull, E. O. Wollan and W. A. Strauser, *Phys. Rev.*, 81 (1951) 483.
26. L. Néel, *Compt. Rend.*, 230 (1950) 375.
27. P. Weiss, *J. Phys. (Paris)*, 8 (1899) 542; 4 (1905) 469, 829; P. Weiss and R. Forrer, *Ann. Phys. (Paris)*, 12 (1929) 279.
28. F. Bertaut, *Compt. Rend.*, 234 (1952) 1295.
29. H. Forestier and G. Guiot-Guillain, *Compt. Rend.*, 230 (1950) 1844.
30. G. Guiot-Guillain and H. Forestier, *Compt. Rend.*, 231 (1951) 1832.
31. H. Forestier and G. Guiot-Guillain, *Compt. Rend.*, 235 (1952) 48; G. Guiot-Guillain and H. Forestier, *Compt. Rend.*, 237 (1953) 1654.
32. R. Pauthenet and P. Blum, *Compt. Rend.*, 239 (1954) 33.
33. L. Néel, *Compt. Rend.*, 239 (1954) 8.
34. G. Guiot-Guillain, R. Pauthenet and H. Forestier, *Compt. Rend.*, 239 (1954) 155.
35. F. Bertaut and F. Forrat, *Compt. Rend.*, 242 (1956) 382.
36. R. Pauthenet, *Ann. Phys. (Paris)*, 3 (1958) 424.
37. R. Pauthenet, *Ann. Phys. (Paris)*, 7 (1952) 710.
38. H. A. Kramers, *Physica*, 1 (1934) 182.

the specific heat of nickel, Louis Néel has mainly concentrated on theoretical problems, which have formed the subject of more than 150 publications. Besides his discovery of the concepts of antiferromagnetism and ferrimagnetism and its consequences, for which he was awarded the Nobel Prize, Louis Néel tackled and solved a number of other problems and extended our knowledge of many aspects of magnetism. The most important of these are as follows: theory of Rayleigh's Laws; magnetic properties of fine grains; magnetic viscosity; internal dispersion fields; superantiferromagnetism; and hysteresis.

The following distinctions and honours have been awarded to Professor Néel: Chevalier de la Légion d'Honneur (military) in 1940, Officier in 1951, Commandeur in 1958, and Grand Officier in 1966; Croix de Guerre with palms (1940); Commandeur de l'Ordre des Palmes Académiques (1957); Chevalier du Mérite Social (1963); Holweck Prize (1952); Gold Medal of the Centre National de la Recherche Scientifique (1965). He is a member of the French Academy of Science (Paris, 1953); a foreign member of the Soviet Academy of Science (1959), the Royal Dutch Academy of Science (1959), the Deutsche Akademie der Naturforscher Leopoldina (1964), the Rumanian Academy (1965), the Royal Society (London) (1966), and the American Academy of Arts and Sciences (1966).

Prof. Néel is honorary doctor of the Universities of Graz (1948), Nottingham (1951), Oxford (1958), Louvain (1965), Newcastle (1965), Coimbra (1966), Sherbrooke (1967), and Iassy (1971). He holds an honorary degree from the Polytechnic Institute of Turin (1960). He is an honorary member and former president (1957) of the Société Française de Physique. From 1963 to 1966 he was President of the International Union of Pure and Applied Physics.

Biography

Louis Néel was born in Lyons on 22 November 1904. In 1931 he married Hélène Hourticq; they have three children, Marie Françoise, Attachée d'Administration at the Conseil d'Etat, Marguerite, married to Guély, Professeur agrégée d'histoire, and Pierre, who is a television producer. Louis Néel studied at the Ecole Normal Supérieure in Paris from 1924–1928, where he was appointed lecturer in 1928. In 1932 he obtained the degree of Doctor of Science at the University of Strasbourg, where he was appointed Professor at the Faculty of Science (1937–1945). He was Professor in Grenoble since 1945. In 1946 he became Director of the laboratory for electrostatics and metal physics (Centre National de la Recherche Scientifique). From 1954 until 1970 he was Director of the Institut Polytechnique de Grenoble and of the Ecole Française de Papeterie; in 1970 he was appointed President of the Institut National Polytechnique in Grenoble. He served as director of the Centre d'Etudes nucléaires de Grenoble from 1956 to 1970. From 1949 to 1969 he was a member of the Board of Directors of the C.N.R.S.; scientific adviser to the French Navy since 1952; French representative at the Scientific Committee of the North Atlantic Treaty Organization.

Louis Néel began his first research work on magnetism between 1928 and 1939 in Professor Weiss' laboratory in Strasbourg. Called up for war service in 1939, he worked on the defence of ships of the French fleet against German magnetic mines and invented an effective new method of protection (neutralization). After the Armistice of 1940, he went to Grenoble and established the Laboratoire d'Electrostatique et de Physique du Métal, which in 1946 became one of the external laboratories of the Centre National de la Recherche Scientifique. This laboratory extended rapidly and gave rise to new laboratories; even so, it still has a staff of more than 250 at the present time.

In 1956 Louis Néel created and subsequently developed, as part of the French Atomic Energy Commission, the Centre d'Etudes Nucléaires de Grenoble. He also contributed to the decision to install the Franco-German high-flux reactor in Grenoble (1967).

Although he continued with research, sometimes critical and difficult, on

Name Index

Alfvén, H. 301–305, *306–317*
Alfvén, J. 317
Allen, C.W. 218
Alston, M. 279
Alvarez, L.W. 237–240, *241–292*
Amaldi, E. 43
Ampère, A.M. 304
Anderson, C.D. 241, 243, 244
Anderson, H.L. 242
Anderson, J.A. 284
Atkinson, R. d'E. 213, 216
Bacher, 43
Bär, R. 186
Basov, G.F. 106
Basov, N.G. 53–57, 61, 69, *89–104*, *106–109*, 118
Basova, K.T. 109
Becquerel, A.H. 215
Belenov, E.M. 108
Bennett, 74
Bergman, I. 149
Bertaut, F. 335, 336
Bethe, H.A. 20, 43, 124, 135, 170–172, 209–214, *215–236*, 244, 245
Bhagavantam, S. 187
Birmingham, B. 257
Bitter, F. 184, 187, 206, 325
Bizette, H. 325
Blamont, 190–192
Bloch, E. 61, 127, 129, 131, 186, 205, 291
Bloembergen, N. 68, 69
Blum, P. 336
Blumberg, R. 254, 255
Bode, 8
Bogdankevich, O.V. 107, 108
Bohr, A.B. 49
Bohr, N. 4, 40, 43–48, 140, 158, 205, 235, 304
Boltzmann, L. 218, 220
Born, M. 38
Bose, 144
Bouchiat-Guiochon, M.A. 195
Bradner, H. 262, 269
Braunbeck, 130
Bricard, J. 206
Brillouin, L. 83
Brobeck, W. 246
Brossel, J. 184, 185, 187–193, 206
Brown, F.H. 88
Brown, R.M. 244
Brueckner, K.A. 20, 46, 242
Bruhat, G. 205
Brunold, 205
Burkhard, J. 264
Butler, C.C. 244, 295
Byrns, R. 255
Cabannes, J. 186
Cagnac, 193, 195, 196
Carlson, A.G. *see* Ekspong, A.G.
Carrol, C. 153
Cassen, B. 242
Caughlan, G.R. 224, 226
Chadwick, J. 40, 41, 44, 241, 243
Chelton, D. 257
Chew, G.F. 284
Cohen-Tannoudji, J. 185, 194, 197–201
Collins, 73
Condon, E.U. 196, 216, 242
Conversi, M. 241, 269
Cornog, 291
Cosset, E. 207
Crawford Jr., F.S. 246, 249, 251, 254, 272

Crawford, J.A. 270
Critchfield, C.L. 221
Curie, M. 215
Curie, P. 215
Dalitz, R.H. 278
Dancoff, S.M. 130–135, 242
Danilychev, V.A. 108
Danos, 48
Darwin, C. 215
Daure, P. 205, 206
Davey, P.G. 263
Davies Jr., R. 227
Dehmelt, H.G. 196, 197
Demarque, P.R. 219
Descoubes, J.P. 193
De Shalit, A. 49
Devyatkov, A.N. 85, 107
Dicke, R.H. 90
Dirac, P.A.M. 110, 124, 126–129, 135, 140, 144, 147, 155, 156, 165, 168, 172, 177
Dow, J.M. 300
Dumke, W.P. 96
Eberhard, P. 274, 279, 281
Eckman, G. 254, 262
Eddington, 212, 216, 218, 219
Edlén, B. 55
Einstein, A. 10, 12, 55, 59, 93, 110, 140, 144, 158, 239
Ekspong, A.G. 242
Elliot, 49
Elsasser, W. 4, 21, 29, 46
Epstein, E. 135
Esclangon, F. 187
Ewald, P.P. 236
Fälthammar, C.-G. 317
Faraday, M. 176
Faroux, J.P. 196
Feher, 70
Fermi, E. 29, 39, 43–45, 144, 188, 242
Ferro-Luzzi, M. 286
Feynman, R.P. 121–125, 135, 136, 140, 148, *155–179*
Fierz, 130, 131, 135
Fitch, V. 246
Flowers, 49
Fock, V.A. 12, 117, 127
Foley, 125
Forestier, H. 336
Forrat, F. 337
Fowler, W.A. 223–225
Fowler, W.B. 243
Franck, J. 38, 188, 196, 259, 262, 263
Frank, F.C. 272
Franzen, W. 195
Frenkel, 161
Friesen, S. von, *see* Von Friesen, S.
Frisch, R. 187
Frish, S.E. 117
Fröhlich, 205
Gamow, G. 41, 46, 213, 216
Gardner, E. 242
Garnjost, M., *see* Alston, M.
Gell-Mann, M. 16, 246, 274, 276, 286, 287, 293–298, *300*
Glaser, D. 240, 248, 249, 251
Goeppert, F. 38
Goeppert Mayer, M. 1–5, *20–39*, 47, 48
Goldenberg, 67
Goldhaber, M. 48
Goldhaber, S. 246
Goldschmidt, V.M. 46
Good, M.L. 246, 273, 279
Gordon, J.P. 61, 63, 66
Gorter, C.J. 324
Gorter, 183, 206
Gow, J.D. 254, 257–259, 261, 269
Grasjuk, A.Z. 107
Graziano, W. 277–279
Greiner, E. 205
Gross, E.K. 117
Grossetête, F. 196
Guggenheim, 46
Guiochon, M.A. 190–192
Guiochon, M.A., *see also* Bouchiat-Guiochon, M.A.
Guiot-Guillain, G. 336
Gurney, R.W. 216
Gustafsson, T. 303
Hall, R.N. 75

Serres, A. 328
Shalit, A. de, *see* De Shalit, A.
Sheglov, V.A. 108
Shelepina, G.A. 119
Shull, C.G. 325, 332
Shutt, R.P. 243
Skinner, H.W.B. 202
Sklizkov, G.V. 107, 108
Skobeltzyn, D.V. 118
Slotnick, 173, 174
Smart, S.J. 325
Snyder, J. 263, 265
Solmitz, F.T. 264, 265
Sommerfeld, A. 186, 205, 235
Squire, C. 325
Stefan, 218
Steinberger, J. 242
Steinwedel, 48
Steller, J. 242
Stevenson, E.C. 241
Stevenson, M.L. 246, 249, 251, 254, 266, 281
Strakhovsky, G.M. 106
Strauss, L. 258
Street, J.C. 241
Strömgren, 218, 220
Suess, H.E. 4, 21, 29, 36, 39, 46, 47
Taft, H. 265
Talmi, I. 36
Teller, E. 21, 39, 48, 216, 270
Terhune, R.W. 77
Thompson, R.W. 245
Ticho, H.K. 270, 273, 274, 279, 286
Tomonaga, H. 137
Tomonaga, S. 121–125, *126–139*, 148
Townes, C.H. 53–57, *58–88*
Townes, E.H. 87
Townes, H.K. 87
Tripp, R.D. 268, 286
Tsaï, B. 325
Urey, H. 39
Van Vleck, J.H. 325
Veksler, V.I. 117
Verwey, E.J.W. 327, 328
Villain, 339
Virge, R.W. 246
Vitkevitch, V.V. 118
Vleck, J.H. van, *see* Van Vleck, J.H.
Von Friesen, S. 239, 246
Von Laue, M. 41
Von Neumann, J. 258
Von Weizsäcker, C.F. 8, 213, 216, 221
Vul, B.M. 107
Walker, W.D. 184
Waller, C. 242
Waller, I. 3, 123, 183, 295
Watson, M.B. 286
Watt, R. 254, 261, 262
Weber, J. 61
Weil, L. 327
Weinmann, 130
Weiss, P. 304, 305, 318–321, 334, 342
Weisskopf, V. 46, 129, 131, 132, 275
Weizsäcker, C.F. von, *see* Von Weizsäcker, C.F.
Wheeler, A. 18, 44, 158, 159, 163, 172
Wiens, 291
Wigner, D. 18
Wigner, E.P. 1–5, *6–15*, *18–19*, 27, 37, 42, 43, 45, 47, 48
Wigner, M. 18
Wilson, 248
Windaus, 38
Winter, J.M. 184, 194
Wojcicki, S.G. 277–279
Wolff, M. 38
Wood, J.G. 249–251
Woodbury, E.J. 76
Wu, C.S. 272
Yang, C.N. 4, 6, 10, 15, 246, 272
York, H.F. 242
Yoshimori, 339
Yukawa, H. 45, 128, 137, 241, 242, 295
Zakharov, S.D. 108
Zeiger, H.J. 61, 63, 66
Zel'dovitch, Ya.B. 272
Zimmerman, B.A. 224
Zuev, V.S. 106, 107

Molchanova, Z.A. 106
Morozov, V.N. 108
Motley, R. 246
Mottelson, 49
Moyer, B.J. 242
Murray, J.J. 274
Nagle, D.E. 249, 251
Nahashima, Y. 139
Nakano, 296
Nasibov, A.S. 108
Neddermeyer, S.H. 241
Néel, L. 207, 301–305, *318–343*
Ne'eman, Y. 286, 287, 297
Neumann, J. von, *see* Von Neumann, J.
Newton, I. 6–8, 11
Ng, W.K. 76
Nikitin, V.V. 106, 108
Nilsson, 49
Nishijima, K. 246, 276, 296
Nishina, Y. 126, 127, 137
Nordsieck, 130, 131
Norgren, D. 260
O'Ceallaigh, C. 245
Oliphant, M.L.E. 278
Omont, A. 196
Oppenheimer, J.R. 153
Oppenheimer, R. 48
Oraevsky, A.N. 106, 108
Pais, A. 132, 243, 246, 295
Pancini, E. 241
Pankratov, A.V. 109
Panofsky, W.K.H. 268, 269
Papalekski, I.D. 117
Papalekski, N.D. 117
Parmentier, D. 251–253
Pauli, W. 44, 45, 124, 126, 130, 131, 135, 140, 145, 242
Pauthenet, E. 336, 338
Pebay-Peyroula, J.C. 193
Peierls, R.F. 235
Percy, J.R. 219
Perlman, 24
Perrier, C. 285
Pevsner, A. 285
Piccioni, O. 241
Pitzer, 291
Placzek, 186
Planck, M.K.E.L. 58
Podolsky, W.J. 127
Popov, Yu.M. 107, 108
Pound, R.V. 60
Powell, 246, 248, 272
Powell, C.F. 242
Powell, W. 254, 255, 268
Prochorov, A.M. 53–57, 61, 69, 104, 106, 109, *110–119*
Prout, W. 40
Purcell, E.M. 60
Rabi, I.I. 153, 183, 187, 206
Raman, C.V. 186, 187
Ramsey, N.F. 67
Rasetti, F. 188
Reeves, H. 225
Retherford, R.C. 60, 123, 187
Rilke, R.M. 50
Rinta, 254
Rochester, G.D. 244, 295
Rollet, N. 190, 191
Romanus, A.-C. 317
Rosenfeld, A.H. 127, 265, 266, 285
Ross, R.R. 263, 266
Rubinowicz, A. 186, 205
Russell, 217
Rutherford, E. 41, 211, 278
Rytov, S.M. 117
Sakata, 132, 296
Salpeter, E.E. 217, 231, 233
Schawlow, A.L. 56, 66, 70, 88
Schrödinger, E. 41, 44, 45, 167
Schwarzschild, M. 228
Schwemin, A.J. 251–254, 256, 263
Schwinger, J. 16, 121–125, 135, 136, *140–154*, 179
Scovil, H.E.D. 70, 116
Seaborg, G.T. 24, 46
Segrè, E. 243, 285
Seidel, H. 71
Sekiguchi, K. 138
Selivanenko, A.C. 118
Senatsky, Yu.V. 107, 108

Hanle, W. 186
Harkins, 41
Haroche, S. 194, 201
Haxel, O. 4, 29, 39, 46, 47, 51
Heilmann, E.L. 327, 328
Heisenberg, W. 41, 42, 44, 45, 124, 126, 130, 137, 140, 167, 304, 319
Heitler, W. 155, 236
Helmholtz, H.L.F. 215
Hensley, D.C. 225
Hernandez, P. 254, 257–261
Herriott, D. 74
Hertz, G. 58, 188
Hertzsprung, 217
Herzfeld, K.F. 38, 39
Hildebrand, R.H. 249, 251
Hofstadter, R. 42, 281
Holweck, F. 207
Hooper, J.E. 242
Hourticq, H. 342
Houtermans, F.G. 216
Howarth, G. 179
Hoyle, F. 231, 232
Hulsizer, R.I. 263
Humphrey, W.E. 263
Hund, F. 43
Iben Jr., I. 228, 229
Ivanenko, 46
Javan, A. 74–76, 107
Jehle, H. 165, 171
Jensen, J. 51
Jensen, J.H.D. 1–5, 20, 21, 29, 39, *40–51*
Jensen, P. 48
Jordan, P. 45
Kastler, A. 181–185, *186–207*
Kastler, C.-Y. 207
Kastler, D. 207
Kastler, M. 207
Katulin, V.A. 107
Kenney, R.W. 253
Kepler, 6–8
Kerstin, M.E. 317
King, D.T. 242
Klein, O. 45, 211
Kleppner, 67
Kramers, H.A. 339
Kriukov, P.G. 107, 108
Krokhin, O.N. 107, 108
Kurath, D. 49
Kusch, 124
Lamb Jr., W.E. 60, 123, 179, 187
Lattes, C.M.G. 242
Laue, M. von, *see* Von Laue, M.
Lawrence, E. 242, 258, 284, 285
Lee, T.D. 4, 6, 15, 246, 272
Lehmann, J.C. 195
Leighton, R.B. 244
Lenz, W. 50
Leontovich, M.A. 106, 117, 118
Leprince-Ringuet, L. 244
Letokhov, V.S. 107, 108
Lewis, 135
Libby, W. 258
Lloyd, J. 263
Long, E.A. 249
Lorentz, H.A. 129, 157
Low, 284
Lynch, G. 264, 267
McCormick, B. 263
McCullough, 177
McMillan, 242
Maglić, B. 281–283
Mahuet, 205
Maiman, T.H. 73
Mandelshtam, S.L. 108
Manenkov, A.A. 118
Mang, 41
Mann, D. 257
Margerie, J. 201
Mark, J. 255
Markin, E.P. 108
Marshak, R. 245
Massey, H.S.W. 308
Maxwell, J.C. 58, 176
Mayer, J.E. 38
Mayer, M. Goeppert, *see Goeppert Mayer, M.*
Melvin, M.A. 12
Menon, M.G.K. 245

Subject Index

Atomic nucleus, the shell model 20
Coherent radiation, production by atoms and molecules 58
Currents, symmetry and, in particle physics 299
Development of quantum electrodynamics 126
Development of the space–time view of quantum electrodynamics 155
Electrodynamics, quantum, development of 126
Electrodynamics, quantum, development of the space–time view of 155
Electronics, quantum 110
Energy production in stars 215
Events, laws of nature, and invariance principles 6
Glimpses at the history of the nuclear structure theory 40
Hertzian resonances, optical methods for studying 186
Invariance principles, events, laws of nature and 6
Lasers, semiconductor 89
Laws of nature, events, and invariance principles 6
Local molecular field, magnetism and the 318
Magnetism and the local molecular field 318
Molecular field, the local, magnetism and 318
Nature, laws of, events, and invariance principles 6
Nuclear structure theory, glimpses at the history of the 40
Optical methods for studying Hertzian resonances 186
Particle physics, recent developments in 241
Particle physics, symmetry and currents in 299
Plasma physics, space research and the origin of the solar system 306
Production of coherent radiation by atoms and molecules 58
Quantum electrodynamics, development of 126
Quantum electrodynamics, development of the space–time view of 155
Quantum electronics 110
Quantum field theory, relativistic 140
Radiation, coherent, production by atoms and molecules 58
Recent developments in particle physics 241
Relativistic quantum field theory 140
Semiconductor lasers 89
Shell model, the 20
Solar system, the origin of the 306
Space research, plasma physics and the origin of the solar system 306
Stars, energy production in 215
Symmetry and currents in particle physics 299

Index of Biographies

Alfvén, Hannes 317
Alvarez, Luis W. 291
Basov, Nikolai G. 106
Bethe, Hans Albrecht 235
Feynman, Richard P. 155
Gell-Mann, Murray 300
Goeppert Mayer, Maria 38
Jensen, J. Hans D. 51
Kastler, Alfred 205
Mayer, Maria Goeppert 38
Néel, Louis 342
Prochorov, Alexander M. 117
Schwinger, Julian 153
Tomonaga, Sin-itiro 126
Townes, Charles H. 87
Wigner, Eugene P. 6