NOBEL LECTURES

PHYSICS

1942 – 1962

World Scientific

NOBEL LECTURES IN PHYSICS

1942–1962

NOBEL LECTURES

INCLUDING PRESENTATION SPEECHES
AND LAUREATES' BIOGRAPHIES

PHYSICS

CHEMISTRY

PHYSIOLOGY OR MEDICINE

LITERATURE

PEACE

ECONOMIC SCIENCES

NOBEL LECTURES

INCLUDING PRESENTATION SPEECHES
AND LAUREATES' BIOGRAPHIES

PHYSICS

1942–1962

World Scientific
Singapore • New Jersey • London • Hong Kong

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 (1942–1962)

All rights reserved.

ISBN 981-02-3403-1

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

Physics 1942

Prize not awarded.

Physics 1943 and 1944

OTTO STERN

[1943]

«for his contribution to the development of the molecular ray method and his discovery of the magnetic moment of the proton»

ISIDOR ISAAC RABI

[1944]

«for his resonance method for recording the magnetic properties of atomic nuclei»

Physics 1943 and 1944

The following account of Stern's and Rabi's works is by Professor E. Hulthén, Stockholm (Broadcast lecture, 10th December, 1944)

There is a certain relation between electric and magnetic phenomena in that the magnetic field can generally be ascribed to the presence of electric currents. It was in this way that the famous Ampère sought to trace magnetism back to rotary currents of electricity in the particles of matter, the atoms and molecules. This hypothesis has in fact been confirmed, *inter alia* by spectroscopical investigations into light sources placed in very strong magnetic fields. However, certain difficulties arose when it came to accounting in detail for the influence of the magnetic field on the movement of electrons, which here represents the electric currents in the interior of the atom. For the electrons proved disinclined to obey the electrodynamic laws which have otherwise so brilliantly demonstrated their validity in, for instance, the field of electrotechnics. *Inter alia*, it seemed as if the small, freely moving atomic magnet in the source of light was only capable of assuming certain discrete local positions in relation to the direction of the applied field. I shall start, then, with a reference to an experiment which for the first time revealed this remarkable so-called directional or space-quantization effect.

The experiment was carried out in Frankfurt in 1920 by Otto Stern and Walter Gerlach, and was arranged as follows: In a small electrically heated furnace, was bored a tiny hole, through which the vapour flowed into a high vacuum so as to form thereby an extremely thin beam of vapour. The molecules in this so-called atomic or molecular beam all fly forwards in the same direction without any appreciable collisions with one another, and they were registered by means of a detector, the design of which there is unfortunately no time to describe here. On its way between the furnace and the detector the beam is affected by a non-homogeneous magnetic field, so that the atoms – if they really *are* magnetic – become unlinked in one direction or another, according to the position which their magnetic axles may assume in relation to the field. The classical conception was that the thin and clear-cut beam would consequently expand into a diffuse beam, but in actual fact the opposite proved to be the case. The two experimenters found that the beam divided up into a number of relatively still sharply defined beams, each

corresponding to one of the just mentioned discrete positional directions of the atoms in relation to the field. This confirmed the space-quantization hypothesis. Moreover, the experiment rendered it possible to estimate the magnetic factors of the electron, which proved to be in close accord with the universal magnetic unit, the so-called «Bohr's magneton».

When Stern had, so to speak, become his own master, having been appointed Head of the Physical Laboratory at Hamburg in 1923, he was able to devote all his energies to perfecting the molecular beam method. Among many other problems investigated there was a particular one which excited considerable interest.

It had already been realized when studying the fine structure of the spectral lines that the actual nucleus of the atom, like the electron, possesses a rotation of its own, a so-called «spin». Owing to the minute size of the nuclear magnet, estimated to be a couple of thousand times smaller than that of the electron, the spectroscopists could only determine its size by devious ways – and that too only very approximately. The immense interest attaching in this connection to a determination of the magnetic factors of the hydrogen nucleus, the so-called proton, was due to the fact that the proton, together with the recently discovered neutron, forms the basic constituent of all the elements of matter; and if these two kinds of particles were to be regarded, like the electron, as true elementary particles, indivisible and uncompounded, then as far as the proton is concerned, its magnetic factor would be as many times smaller than the electron's as its mass is greater than the electron's, implying that the magnetic factor of the proton must be, in round figures, 1,850 times smaller than the electron's. Naturally then, it aroused great interest when, in 1933, Stern and his colleagues made this determination according to the molecular beam method, it being found that the proton factor was about $2\frac{1}{2}$ times greater than had theoretically been anticipated.

Let us now for a moment touch upon Rabi's achievements in this field. Returning to the essential point of the problem, let us put the question: How does the atom react to the magnetic field? According to a theorem stated by the English mathematician Larmor, this influence may be ascribed to a relatively slow precession movement on the part of the electron and the atomic nucleus around the field direction – a gyromagnetic effect most closely recalling the gyroscopic movement performed by a top when it spins around the vertical line. If the strength of the magnetic field is known, the magnetic factor of the electron and of the atomic nucleus can also be estimated by this means, provided that we can observe and measure these precessional frequen-

cies. Rabi solved the problem in a manner as simple as it was brilliant. Within the magnetic field was inserted a loop of wire, attached to an oscillating circuit the frequency of which could be varied in the same manner as we tune in our radio receiving set to a given wavelength. Now, when the atomic beam passes through the magnetic field, the atoms are only influenced on condition that they precess in time with the electric current in the oscillating circuit. This influence might perhaps be described graphically: the nucleus performs a vault (salto) – the technical term for which is a « quantum jump » – thereby landing in another positional direction to the field. But this means that the atom has lost all chance of reaching the detector and of being registered by it. The effect of these quantum jumps is observable by the fact that the detector registers a marked resonance minimum, the frequency position of the registration being determined with the extraordinary precision achievable with the radio frequency gauge. By this method Rabi has literally established radio relations with the most subtile particles of matter, with the world of the electron and of the atomic nucleus.

OTTO STERN

The method of molecular rays

Nobel Lecture, December 12, 1946

In the following lecture I shall try to analyze the method of molecular rays. My aim is to bring out its distinctive features, the points where it is different from other methods used in physics, for what kind of problems it is especially suited and why. Let me state from the beginning that I consider the directness and simplicity as the distinguishing properties of the molecular ray method. For this reason it is particularly well suited for shedding light on fundamental problems. I hope to make this clear by discussing the actual experiments.

Let us first consider the group of experiments which prove directly the fundamental assumptions of the kinetic theory. The existence of molecular rays in itself, the possibility of producing molecular rays, is a direct proof of one fundamental assumption of that theory. This assumption is that in gases the molecules move in straight lines until they collide with other molecules or the walls of the containing vessel. The usual arrangement for producing molecular rays is as follows (Fig. 1): We have a vessel filled with gas or vapor, the oven. This vessel is closed except for a narrow slit, the oven slit. Through this slit the molecules escape into the surrounding larger vessel which is continually evacuated so that the escaping molecules do not suffer any collisions. Now we have another narrow slit, the collimating slit, opposite and parallel to the oven slit. If the molecules really move in straight lines then the collimating slit should cut out a narrow beam whose cross section by simple geometry can be calculated from the dimensions of the slits and their distance. That it is actually the case was proven first by Dunoyer in 1911. He used sodium vapor and condensed the beam molecules hitting the wall by cooling it with liquid air. The sodium deposit formed on the wall had exactly the shape calculated under the assumption that the molecules move in straight lines like rays of light. Therefore we call such a beam a «molecular ray» or «molecular beam».

The next step was the direct measurement of the velocity of the molecules. The kinetic theory gives quite definite numerical values for this velocity, depending on the temperature and the molecular weight. For example, for silver atoms of 1,000° the average velocity is about 600 m/sec (silver mole-

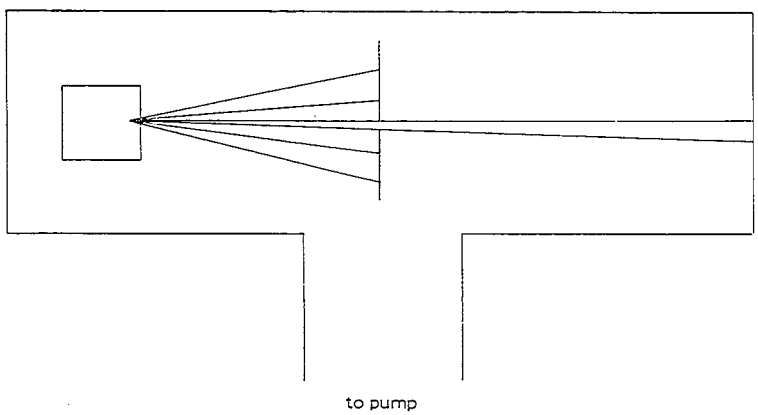


Fig. 1. Arrangement for producing molecular rays.

cules are monoatomic). We measured the velocity in different ways. One way – historically not the first one – was sending the molecular ray through a system of rotating tooth wheels, the method used by Fizeau to measure the velocity of light. We had two tooth wheels sitting on the same axis at a distance of several cm. When the wheels were at rest the molecular beam went through two corresponding gaps of the first and the second wheel. When the wheels rotated a molecule going through a gap in the first wheel could not go through the corresponding gap in the second wheel. The gap had moved during the time in which the molecule travelled from the first wheel to the second. However, under a certain condition the molecule could go through the next gap of the second wheel, the condition being that the travelling time for the molecule is just the time for the wheel to turn the distance between two neighboring gaps. By determining this time, that means the number of rotations per second for which the beam goes through both tooth wheels, we measure the velocity of the molecules. We found agreement with the theory with regard to the numerical values and to the velocity distribution according to Maxwell's law.

This method has the advantage of producing a beam of molecules with nearly uniform velocity. However, it is not very accurate.

As the last one in this group of experiments I want to report on experiments carried out in Pittsburgh by Drs. Estermann, Simpson, and myself before the War, which are now being published. In these experiments we used the free fall of molecules to measure their velocities.

In vacuo all bodies, large and small, fall equal distances in equal times, $s =$

$\frac{1}{2} gt^2$ (t = time, s = distance of fall, g = acceleration of gravity). We used a beam of cesium atoms about 2 m long. Since the average velocity of the atoms is about 200 m/sec the travel time is about 1/100 sec. During this time a body falls not quite a mm. So our cesium atoms did not travel exactly on the straight horizontal line through oven and collimating slit but arrived a little lower depending on their velocity. The fast ones fell less, the slow ones more. So by measuring the intensity (the number of cesium atoms arriving per second) at the end of the beam perpendicular to it as a function of the distance from the straight line, we get directly the distribution of velocities (Fig. 2). As you see the agreement with Maxwell's law is very good. I might mention that we measured the intensity not by condensation but by the so-called Taylor-Langmuir method worked out by Taylor in our Hamburg

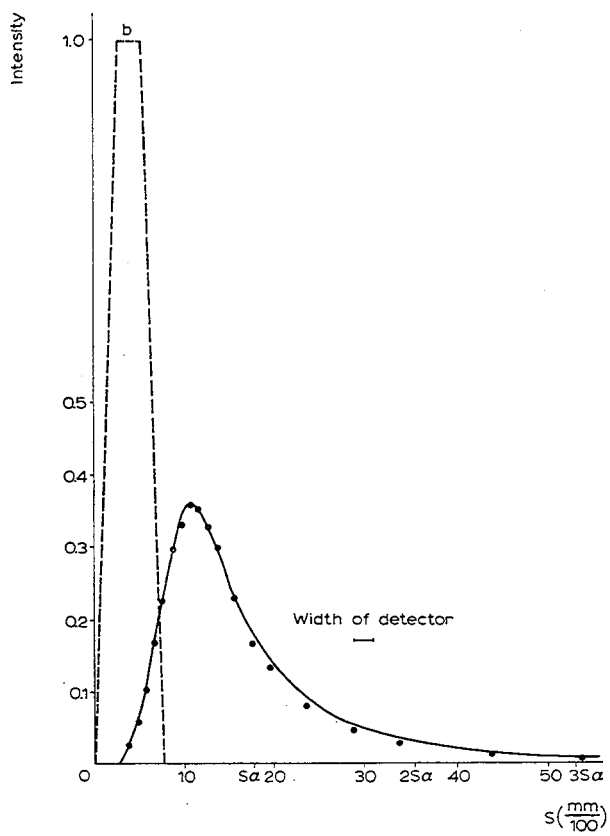


Fig. 2. Gravity deflection of a cesium beam. (*Full line*): calculated from Maxwell's law; (*points*): measurements; (*pecked line b*): undeflected beam.

laboratory in 1928. It is based on Langmuir's discovery that every alkali atom striking the surface of a hot tungsten wire (eventually oxygen-coated) goes away as an ion. By measuring the ion current outgoing from the wire we measured directly the number of atoms striking the wire.

What can we conclude about the method of molecular rays from the group of experiments we have considered so far? It gives us certainly a great satisfaction to demonstrate in such a simple direct manner the fundamentals of the kinetic theory. Furthermore, even if so many conclusions of the theory were checked by experiments that practically no reasonable doubt about the correctness of this part of the theory was possible, these experiments reinforced and strengthened the fundamentals beyond any doubt.

I said this part of the theory.

The classical theory is a grandiose conception. The same fundamental laws govern the movements of the stars, the fall of this piece of chalk, and the fall of molecules. But it turned out that the extrapolation to the molecules did not hold in some respects. The theory had to be changed in order to describe the laws governing the movements of the molecules and even more of the electrons. And it was at this point that the molecular ray method proved its value. Here the experiment did not just check the results of the theory on which there was practically no doubt anyway, but gave a decisive answer in cases where the theory was uncertain and even gave contradictory answers.

The best example is the experiment which Gerlach and I performed in 1922. It was known from spectroscopic experiments (Zeeman effect) that the atoms of hydrogen, the alkali metals, silver, and so on, were small magnets. Let us consider the case of the hydrogen atom as the simplest one even if our experiments were performed with silver atoms. There is no essential difference, and the results were checked with hydrogen atoms a few years later by one of my students in our Hamburg laboratory.

The essential point is that the classical theory and the quantum theory predict quite differently the behavior of the atomic magnets in a magnetic field. The classical theory predicts that the atomic magnets assume all possible directions with respect to the direction of the magnetic field. On the other hand, the quantum theory predicts that we shall find only two directions parallel and antiparallel to the field (new theory, the old one gave also the direction perpendicular to the field).

The contradiction I spoke of is this. At this time according to Bohr's theory one assumed that the magnetic moment of the hydrogen atom is produced by the movement of the electron around the nucleus in the same way

as a circular current in a wire is equivalent to a magnet. Then the statement of the quantum theory means that the electrons of all hydrogen atoms move in planes perpendicular to the direction of the magnetic field. In this case one should find optically a strong double refraction which was certainly not true. So there was a serious dilemma.

Our molecular ray experiment gave a definite answer. We sent a beam of silver atoms through a very inhomogeneous magnetic field. In such a field the magnets are deflected because the field strength on the place of one pole of the magnet is a little different from the field strength acting on the other pole. So in the balance a force is acting on the atom and it is deflected. A simple calculation shows that from the classical theory follows that we should find a broadening of the beam with the maximum intensity on the place of the beam without field. However, from the quantum theory follows that we should find there no intensity at all, and deflected molecules on both sides. The beam should split up in two beams corresponding to the two orientations of the magnet. The experiment decided in favor of the quantum theory (Fig. 3).

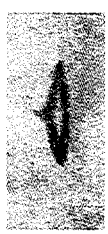


Fig. 3. Discrete beams of deflected molecules.

The contradiction with respect to the double refraction was solved about four years later through the new quantum mechanics in connection with the Uhlenbeck-Goudsmit hypothesis that the electron itself is a small magnet like a rotating charged sphere. But even before this explanation was given, the experiment verified directly the splitting in discrete beams as predicted by the quantum theory.

So again the directness stands out as characteristic for the molecular ray method. However, we can recognize another feature as essential in this experiment, namely that our measuring tool is a macroscopic one. I want to make this point clearer.

The first experiment which gave a direct proof of the fundamental hypoth-

esis of the quantum theory was the celebrated experiment of Franck and Hertz. These workers proved that the energy of one atom can be changed only by finite amounts. By bombarding mercury atoms with electrons they found that the electrons did lose energy only if their energy was higher than 4.7 eV. So they demonstrated directly that the energy of a mercury atom cannot be changed continuously but only by finite amounts, quanta of energy. As a tool for measuring the energy changes of the atom they used electrons, that means an atomic tool. In our experiment we used an electromagnet and slits, that means the same kind of tools we could use to measure the magnetic moment of an ordinary macroscopic magnet. Our experiment demonstrated in a special case a fact, which became fundamental for the new quantum mechanics, that the result of our measurements depends in a characteristic manner on the dimensions of the measured object and that quantum effects become perceptible when we make the object smaller and smaller.

We can see this better when we first consider a group of experiments which demonstrate the wave properties of rays of matter. In his famous theory which became the basis of the modern quantum theory, De Broglie stated that moving particles should also show properties of waves. The wavelength of these waves is given by the equation $\lambda = h/mv$ (h = Planck's constant; m = mass; v = velocity of the particle). The experimental proof was first given in 1927 by Davisson and Germer, and by Thomson for electrons. Some years later we succeeded in performing similar experiments with molecular rays of helium atoms and hydrogen molecules using the cleavage surface of a lithium fluoride crystal as diffraction grating. We could check the diffraction in every detail. The most convincing experiment is perhaps the one where we sent a beam of helium gas through the two rotating tooth wheels which I mentioned at the beginning, thus determining the velocity v in a primitive, purely mechanical, manner. The helium beam then impinged on the lithium fluoride crystal and by measuring the angle between the diffracted and the direct beam we determined the wavelength since we know the lattice constant of the lithium fluoride. We found agreement with De Broglie's formula within the accuracy of our experiments (about 2%). There is no doubt that these experiments could be carried out also by using a ruled grating instead of the crystal. In fact we found hints of a diffracted beam with a ruled grating already in 1928 and with the improved technique of today the experiment would not be too difficult.

With respect to the differences between the experiments with electrons and molecular rays, one can say that the molecular ray experiments go far-

ther. Also the mass of the moving particle is varied (He, H_2). But the main point is again that we work in such a direct primitive manner with neutral particles.

These experiments demonstrate clearly and directly the fundamental fact of the dual nature of rays of matter. It is no accident that in the development of the theory the molecular ray experiments played an important role. Not only the actual experiments were used, but also molecular ray experiments carried out only in thought. Bohr, Heisenberg, and Pauli used them in making clear their points on this direct simple example of an experiment. I want to mention only one consideration concerning the magnetic deflection experiment because it shows the fundamental limits of the experimental method.

First, it is clear that we cannot use too narrow slits, otherwise the diffraction on the slit will spread out the beam. This spreading out can roughly be described as the deflection of the molecules by an angle which is the larger the narrower the slit and the larger the De Broglie wavelength is. Therefore it causes a deflection of the molecules proportional to the distance which the molecule has traversed or to the length of the beam or to the time t since the molecule started from the collimating slit. The deflection by the magnetic force must be appreciably larger if we want to measure the magnetic moment. Fortunately this deflection is proportional to the square of the length of the beam or the time t, essentially as in the case of the gravity ($s = \frac{1}{2} gt^2$). Consequently it is always possible, by making the beam long enough, that means the time t large enough, to make the magnetic deflection larger than the deflection by diffraction. On the other hand it follows that a minimum time is necessary to measure the magnetic moment and this minimum time gets larger when the magnetic deflection, that means the magnetic moment, gets smaller. That is a special case of a general law of the new quantum mechanics. This law – applied to the measurement of moments – says that for *every* method using the same field strength the minimum time is the same. This circumstance was decisive in the group of experiments measuring the magnetic moment of the proton.

The theory predicted that the magnetic moments of electron and proton should be inversely proportional to the masses of those particles. Since the proton has about a two thousand times larger mass than the electron, its magnetic moment should be about two thousand times smaller. Such a small moment could not be measured by the spectroscopic method (Zeeman effect) but we (Frisch, Estermann, and myself) succeeded in measuring it by

using a beam of neutral hydrogen molecules. I do not have time to go into the details. The main point is that in measuring with molecular rays we use a longer time t. In the spectroscopic method this time is the lifetime of the excited atom which emits light by jumping into the normal state. Now this lifetime is generally of the order of magnitude of 10^{-8} sec. Working with molecular rays we use atoms (or molecules) in the normal state whose lifetime is infinite. The duration of our experiment is determined by the time t which the atom (or molecule) spends in the magnetic field. This time was of the order of magnitude of 10^{-4} sec (the length of the field about 10 cm and the velocity of the molecules about 1 km/sec). So our time is about 10,000 times larger and we can measure 10,000 times smaller moments with molecular rays than spectroscopically.

The result of our measurement was very interesting. The magnetic moment of the proton turned out to be about $2\frac{1}{2}$ times larger than the theory predicted. Since the proton is a fundamental particle – all nuclei are built up from protons and neutrons – this result is of great importance. Up to now the theory is not able to explain the result quantitatively.

It might seem now that the great sensitivity as shown in the last experiment is also a distinctive and characteristic property of the molecular ray method. However, that is not the case. The reason for the sensitivity as we have seen is that we make our measurements on atoms in the normal state. But of course many of the other experimental methods do that also.

We can see the situation clearly by considering the last achievement of the molecular ray method, the application of the resonance method by Rabi.

With the deflection method it is difficult to measure the moment to better than several per cent, mainly because of the difficulty of measuring the inhomogeneity in such small dimensions. With the resonance method, Rabi's accuracy is better than $1^0/_{00}$, practically the theoretical limit given by the duration of about 10^{-4} sec of the measurement. Theoretically it would be possible to increase the accuracy simply by making this time longer. But that would mean making the beam longer and for practical reasons we cannot go much farther in this direction. In this connection it is significant that perhaps the best new measurements of the magnetic moments of the proton, neutron, and deuteron were made with the resonance method, however not using molecular rays but just liquid water with practically no limit for the duration of the measurement. So the sensitivity cannot be considered as a distinguishing property of the molecular ray method. However, that we have

such clear-cut simple conditions was the reason for applying the ultrasensitive resonance method first to molecular rays.

In conclusion I would like to summarize as follows: The most distinctive characteristic property of the molecular ray method is its simplicity and directness. It enables us to make measurements on isolated neutral atoms or molecules with macroscopic tools. For this reason it is especially valuable for testing and demonstrating directly fundamental assumptions of the theory.

Biography

Otto Stern was born in Sorau, Upper Silesia, Germany, on February 17, 1888. In 1892 he moved with his parents to Breslau, where he attended high school. He began to study physical chemistry in 1906, receiving his Ph.D. degree from the University of Breslau in 1912. In the same year he joined Einstein at the University of Prague and later followed him to the University of Zurich, where he became Privatdocent of Physical Chemistry at the Eidgenössische Technische Hochschule in 1913.

In 1914 he went to the University of Frankfurt am Main as Privatdocent of Theoretical Physics, remaining there until 1921, except for a period of military service. From 1921 to 1922 he was Associate Professor of Theoretical Physics at the University of Rostock, becoming, in 1923, Professor of Physical Chemistry and Director of the laboratory at the University of Hamburg, where he remained until 1933. In that year he moved to the United States, being appointed Research Professor of Physics at the Carnegie Institute of Technology, Pittsburgh where he remained until 1945, then becoming professor emeritus.

His earliest work was in the field of theoretical physics, especially that of statistical thermodynamics and quantum theory, on which he has published important papers. After 1919, his attention was directed more to experimental physics. His development and application of the molecular beam method proved to be a powerful tool for investigating the properties of molecules, atoms and atomic nuclei. One of the early applications of this was the experimental verification of Maxwell's law of velocity distribution in gases. He collaborated with Gerlach to work on the deflection of atoms by the action of magnetic fields on their magnetic moment, then went on to measure the magnetic moments of sub-atomic particles, including the proton. His work on the production of interference by rays of hydrogen and helium was a striking demonstration of the wave nature of atoms and molecules.

Dr. Stern was awarded an LL.D. by the University of California, Berkeley, 1930. He is a member of the National Academy of Sciences (USA), the

American Association for the Advancement of Science, and the Philosophical Society. He holds foreign membership of the Royal Danish Academy of Sciences. He lives at Berkeley, California.

No Lecture was delivered by Professor I.I. Rabi.

Biography

Isidor Isaac Rabi was born in Raymanov, Austria, on July 29, 1898, the son of David Rabi and Janet Teig. He was brought to the United States by his family, in 1899, and his early education was in New York City (Manhattan and Brooklyn). In 1919 he graduated Bachelor of Chemistry at Cornell University (New York). After three years in non-scientific occupation, he started postgraduate studies in physics at Cornell in 1921, which he later continued at Columbia University. In 1927 he received his Ph.D. degree for work on the magnetic properties of crystals. Aided by fellowships, he spent two years in Europe, working at different times with Sommerfeld, Bohr, Pauli, Stern, and Heisenberg. On his return in 1929 he was appointed lecturer in Theoretical Physics at Columbia University, and after promotion through the various grades became professor in 1937.

In 1940 he was granted leave from Columbia to work as Associate Director of the Radiation Laboratory at the Massachusetts Institute of Technology on the development of radar and the atomic bomb. In 1945 he returned to Columbia as executive officer of the Physics Department. In this capacity he is also concerned with the Brookhaven National Laboratory for Atomic Research, Long Island, an organization devoted to research into the peaceful uses of atomic energy.

His early work was concerned with the magnetic properties of crystals. In 1930 he began studying the magnetic properties of atomic nuclei, developing Stern's molecular beam method to great precision, as a tool for measuring these properties. His apparatus was based on the production of ordinary electromagnetic oscillations of the same frequency as that of the Larmor precession of atomic systems in a magnetic field. By an ingenious application of the resonance principle he succeeded in detecting and measuring single states of rotation of atoms and molecules, and in determining the mechanical and magnetic moments of the nuclei.

Prof. Rabi has published his most important papers in *The Physical Review*, of which he was an Associate Editor for two periods. In 1939 he received the Prize of the American Association for the Advancement of Science and, in

1942, the Elliott Cresson Medal of the Franklin Institute. He was awarded the Medal for Merit, the highest civilian award in World War II, in 1948, the King's Medal for Service in the Cause of Freedom the same year, and is an Officer of the Legion of Honour.

He is an honorary D.Sc. of Princeton, Harvard, and Birmingham Universities. He is a Fellow of the American Physical Society (was its President in 1950) and a member of the National Academy of Sciences, the American Philosophical Society, and of the American Academy of Arts and Sciences.

In 1959 he was appointed a member of the Board of Governors of the Weizmann Institute of Science, Rehovoth, Israel. He holds foreign memberships of the Japanese and Brazilian Academies, and is a member of the General Advisory Committee to the Arms Control and Disarmament Agency, and of the United States National Commission for UNESCO. At the International Conference on Peaceful Uses of Atomic Energy (Geneva, 1955) he was the United States delegate and Vice-President. He is also a member of the Science Advisory Committee of the International Atomic Energy Agency.

Dr. Rabi married Helen Newmark in 1926. They have two daughters. His recreations are travel, walking, and the theatre.

Physics 1945

WOLFGANG PAULI

«for the discovery of the Exclusion Principle, also called the Pauli Principle»

Physics 1945

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

Your Majesty, Royal Highnesses, Ladies and Gentlemen.

Rutherford gave an impulse of fundamental importance to research in atomic structure when in 1911 he found that an essential part of an atom is a positively charged nucleus in which practically the whole of the mass of the atom is concentrated, the electrons, which were discovered earlier, being grouped around the nucleus. During the first two decades following on Rutherford's discovery, the attention of most atomic physicists was focussed on the phenomena connected with the electronic configuration. According to the theory established by Bohr in 1913, and afterwards developed by him and by other scientists, we may describe those states of the atom for which its energy has a definite value by saying that each electron revolves around the nucleus. The energy corresponding to an electron orbit is defined by whole numbers, called « quantum numbers », which so to speak enumerate the energy states of the electrons. The famous theory of the atomic structure of all atoms advanced by Bohr in 1921 asserts that the electrons of an atom are arranged in groups which have different mean distances from the nucleus and are each characterized by two quantum numbers.

Important contributions to the solution of the problem of electronic configuration were made in the following years by Landé and Stoner.

At this stage of the development of atomic theory, Wolfgang Pauli made a decisive contribution through his discovery in 1925 of a new law of Nature, the *exclusion principle* or *Pauli principle*. The 1945 Nobel Prize in Physics has been awarded to Pauli for this discovery.

Pauli based his investigation on a profound analysis of the experimental and theoretical knowledge in atomic physics at the time. He found that four quantum numbers are in general needed in order to define the energy state of an electron. He then pronounced his principle, which can be expressed by saying that there cannot be more than one electron in each energy state when this state is completely defined. Three quantum numbers only can be related to the revolution of the electron round the nucleus. The necessity of a fourth quantum number proved the existence of interesting properties of the electron.

Other physicists found that these properties may be interpreted by stating that the electron has a «spin», i.e. that it behaves to some extent as if it were rapidly rotating round an axis through its centre of gravity.

Pauli showed himself that the electronic configuration is made fully intelligible by the exclusion principle, which is therefore essential for the elucidation of the characteristic physical and chemical properties of different elements. Among those important phenomena for the explanation of which the Pauli principle is indispensable, we mention the electric conductivity of metals and the magnetic properties of matter.

In 1925 and 1926 essential progress of another kind was made in the quantum theory, which is the foundation of atomic physics. New and revolutionary methods were developed for the description of the motion of particles. The fundamental importance of Pauli's discovery could now be seen more clearly. His principle proved to be an independent and necessary complement to the new quantum theory. Another way of expressing the principle, simpler and of wider applicability than the original one, was given. In this respect Pauli himself made an important contribution which has also had other far-reaching consequences.

During the last two decades atomic research has been more and more focussed on the properties of the atomic nuclei. In this connection it has been even more fully confirmed than before that the Pauli principle must be characterized as a fundamental law of Nature. The principle, first discovered for electrons, has proved to be valid for the nuclei of hydrogen, called *protons*, and also for the *neutrons* which are formed in many nuclear reactions. The neutrons are particles which have no charge but have approximately the same masses as the protons. According to present views any atomic nucleus consists of protons and neutrons. The Pauli principle is therefore essential for the description of the properties of atomic nuclei.

Pauli occupies a leading position in present theoretical physics. He has made many other important contributions to different branches of his science, among them several to nuclear physics.

The Royal Swedish Academy of Sciences much regrets that Professor Pauli has not had the opportunity of being present on this occasion to receive in person his Nobel Prize. The prize will now instead be delivered to the *chargé d'affaires* of the Legation of the United States of America.

Mr. Ravndal. Permit me to request you to receive on behalf of Professor Pauli the Nobel Prize in Physics from the hands of His Majesty.

WOLFGANG PAULI

Exclusion principle and quantum mechanics

Nobel Lecture, December 13, 1946

The history of the discovery of the « exclusion principle », for which I have received the honor of the Nobel Prize award in the year 1945, goes back to my students days in Munich. While, in school in Vienna, I had already obtained some knowledge of classical physics and the then new Einstein relativity theory, it was at the University of Munich that I was introduced by Sommerfeld to the structure of the atom – somewhat strange from the point of view of classical physics. I was not spared the shock which every physicist, accustomed to the classical way of thinking, experienced when he came to know of Bohr's « basic postulate of quantum theory » for the first time. At that time there were two approaches to the difficult problems connected with the quantum of action. One was an effort to bring abstract order to the new ideas by looking for a key to translate classical mechanics and electrodynamics into quantum language which would form a logical generalization of these. This was the direction which was taken by Bohr's « correspondence principle ». Sommerfeld, however, preferred, in view of the difficulties which blocked the use of the concepts of kinematical models, a direct interpretation, as independent of models as possible, of the laws of spectra in terms of integral numbers, following, as Kepler once did in his investigation of the planetary system, an inner feeling for harmony. Both methods, which did not appear to me irreconcilable, influenced me. The series of whole numbers 2, 8, 18, 32... giving the lengths of the periods in the natural system of chemical elements, was zealously discussed in Munich, including the remark of the Swedish physicist, Rydberg, that these numbers are of the simple form $2\,n^2$, if n takes on all integer values. Sommerfeld tried especially to connect the number 8 and the number of corners of a cube.

A new phase of my scientific life began when I met Niels Bohr personally for the first time. This was in 1922, when he gave a series of guest lectures at Göttingen, in which he reported on his theoretical investigations on the Periodic System of Elements. I shall recall only briefly that the essential progress made by Bohr's considerations at that time was in explaining, by means of the spherically symmetric atomic model, the formation of the intermediate

shells of the atom and the general properties of the rare earths. The question, as to why all electrons for an atom in its ground state were not bound in the innermost shell, had already been emphasized by Bohr as a fundamental problem in his earlier works. In his Göttingen lectures he treated particularly the closing of this innermost K-shell in the helium atom and its essential connection with the two non-combining spectra of helium, the ortho- and para-helium spectra. However, no convincing explanation for this phenomenon could be given on the basis of classical mechanics. It made a strong impression on me that Bohr at that time and in later discussions was looking for a *general* explanation which should hold for the closing of *every* electron shell and in which the number 2 was considered to be as essential as 8 in contrast to Sommerfeld's approach.

Following Bohr's invitation, I went to Copenhagen in the autumn of 1922, where I made a serious effort to explain the so-called « anomalous Zeeman effect », as the spectroscopists called a type of splitting of the spectral lines in a magnetic field which is different from the normal triplet. On the one hand, the anomalous type of splitting exhibited beautiful and simple laws and Landé[1] had already succeeded to find the simpler splitting of the spectroscopic terms from the observed splitting of the lines. The most fundamental of his results thereby was the use of half-integers as magnetic quantum numbers for the doublet-spectra of the alkali metals. On the other hand, the anomalous splitting was hardly understandable from the standpoint of the mechanical model of the atom, since very general assumptions concerning the electron, using classical theory as well as quantum theory, always led to the same triplet. A closer investigation of this problem left me with the feeling that it was even more unapproachable. We know now that at that time one was confronted with two logically different difficulties simultaneously. One was the absence of a general key to translate a given mechanical model into quantum theory which one tried in vain by using classical mechanics to describe the stationary quantum states themselves. The second difficulty was our ignorance concerning the proper classical model itself which could be suited to derive at all an anomalous splitting of spectral lines emitted by an atom in an external magnetic field. It is therefore not surprising that I could not find a satisfactory solution of the problem at that time. I succeeded, however, in generalizing Landé's term analysis for very strong magnetic fields[2], a case which, as a result of the magneto-optic transformation (Paschen-Back effect), is in many respects simpler. This early work

was of decisive importance for the finding of the exclusion principle.

Very soon after my return to the University of Hamburg, in 1923, I gave there my inaugural lecture as *Privatdozent* on the Periodic System of Elements. The contents of this lecture appeared very unsatisfactory to me, since the problem of the closing of the electronic shells had been clarified no further. The only thing that was clear was that a closer relation of this problem to the theory of multiplet structure must exist. I therefore tried to examine again critically the simplest case, the doublet structure of the alkali spectra. According to the point of view then orthodox, which was also taken over by Bohr in his already mentioned lectures in Göttingen, a non-vanishing angular momentum of the atomic core was supposed to be the cause of this doublet structure.

In the autumn of 1924 I published some arguments against this point of view, which I definitely rejected as incorrect and proposed instead of it the assumption of a new quantum theoretic property of the electron, which I called a « two-valuedness not describable classically »[3]. At this time a paper of the English physicist, Stoner, appeared[4] which contained, besides improvements in the classification of electrons in subgroups, the following essential remark: For a given value of the principal quantum number is the number of energy levels of a single electron in the alkali metal spectra in an external magnetic field the same as the number of electrons in the closed shell of the rare gases which corresponds to this principal quantum number.

On the basis of my earlier results on the classification of spectral terms in a strong magnetic field the general formulation of the exclusion principle became clear to me. The fundamental idea can be stated in the following way: The complicated numbers of electrons in closed subgroups are reduced to the simple number *one* if the division of the groups by giving the values of the four quantum numbers of an electron is carried so far that every degeneracy is removed. An entirely non-degenerate energy level is already « closed », if it is occupied by a single electron; states in contradiction with this postulate have to be excluded. The exposition of this general formulation of the exclusion principle was made in Hamburg in the spring of 1925[5], after I was able to verify some additional conclusions concerning the anomalous Zeeman effect of more complicated atoms during a visit to Tübingen with the help of the spectroscopic material assembled there.

With the exception of experts on the classification of spectral terms, the physicists found it difficult to understand the exclusion principle, since no meaning in terms of a model was given to the fourth degree of freedom of

the electron. The gap was filled by Uhlenbeck and Goudsmit's idea of electron spin[6], which made it possible to understand the anomalous Zeeman effect simply by assuming that the spin quantum number of one electron is equal to $\frac{1}{2}$ and that the quotient of the magnetic moment to the mechanical angular moment has for the spin a value twice as large as for the ordinary orbit of the electron. Since that time, the exclusion principle has been closely connected with the idea of spin. Although at first I strongly doubted the correctness of this idea because of its classical-mechanical character, I was finally converted to it by Thomas' calculations[7] on the magnitude of doublet splitting. On the other hand, my earlier doubts as well as the cautious expression « classically non-describable two-valuedness » experienced a certain verification during later developments, since Bohr was able to show on the basis of wave mechanics that the electron spin cannot be measured by classically describable experiments (as, for instance, deflection of molecular beams in external electromagnetic fields) and must therefore be considered as an essentially quantum-mechanical property of the electron[8,9].

The subsequent developments were determined by the occurrence of the new quantum mechanics. In 1925, the same year in which I published my paper on the exclusion principle, De Broglie formulated his idea of matter waves and Heisenberg the new matrix-mechanics, after which in the next year Schrödinger's wave mechanics quickly followed. It is at present unnecessary to stress the importance and the fundamental character of these discoveries, all the more as these physicists have themselves explained, here in Stockholm, the meaning of their leading ideas[10]. Nor does time permit me to illustrate in detail the general epistemological significance of the new discipline of quantum mechanics, which has been done, among others, in a number of articles by Bohr, using hereby the idea of « complementarity » as a new central concept[11]. I shall only recall that the statements of quantum mechanics are dealing only with possibilities, not with actualities. They have the form « This is not possible » or « Either this or that is possible », but they can never say « That will actually happen then and there ». The actual observation appears as an event outside the range of a description by physical laws and brings forth in general a discontinuous selection out of the several possibilities foreseen by the statistical laws of the new theory. Only this renouncement concerning the old claims for an objective description of the physical phenomena, independent of the way in which they are observed, made it possible to reach again the self-consistency of quantum theory, which ac-

general assumptions. I had always the feeling and I still have it today, that this is a deficiency. Of course in the beginning I hoped that the new quantum mechanics, with the help of which it was possible to deduce so many half-empirical formal rules in use at that time, will also rigorously deduce the exclusion principle. Instead of it there was for electrons still an exclusion: not of particular states any longer, but of whole classes of states, namely the exclusion of all classes different from the antisymmetrical one. The impression that the shadow of some incompleteness fell here on the bright light of success of the new quantum mechanics seems to me unavoidable. We shall resume this problem when we discuss relativistic quantum mechanics but wish to give first an account of further results of the application of wave mechanics to systems of several like particles.

In the paper of Heisenberg, which we are discussing, he was also able to give a simple explanation of the existence of the two non-combining spectra of helium which I mentioned in the beginning of this lecture. Indeed, besides the rigorous separation of the wave functions into symmetry classes with respect to space-coordinates and spin indices together, there exists an approximate separation into symmetry classes with respect to space coordinates alone. The latter holds only so long as an interaction between the spin and the orbital motion of the electron can be neglected. In this way the para- and ortho-helium spectra could be interpreted as belonging to the class of symmetrical and antisymmetrical wave functions respectively in the space coordinates alone. It became clear that the energy difference between corresponding levels of the two classes has nothing to do with magnetic interactions but is of a new type of much larger order of magnitude, which one called exchange energy.

Of more fundamental significance is the connection of the symmetry classes with general problems of the statistical theory of heat. As is well known, this theory leads to the result that the entropy of a system is (apart from a constant factor) given by the logarithm of the number of quantum states of the whole system on a so-called energy shell. One might first expect that this number should be equal to the corresponding volume of the multi-dimensional phase space divided by h^f, where h is Planck's constant and f the number of degrees of freedom of the whole system. However, it turned out that for a system of N like particles, one had still to divide this quotient by $N!$ in order to get a value for the entropy in accordance with the usual postulate of homogeneity that the entropy has to be proportional to the mass for a given inner state of the substance. In this way a qualitative distinction between

like and unlike particles was already preconceived in the general statistical mechanics, a distinction which Gibbs tried to express with his concepts of a generic and a specific phase. In the light of the result of wave mechanics concerning the symmetry classes, this division by $N!$, which had caused already much discussion, can easily be interpreted by accepting one of our assumptions II and III, according to both of which only *one* class of symmetry occurs in Nature. The density of quantum states of the whole system then really becomes smaller by a factor $N!$ in comparison with the density which had to be expected according to an assumption of the type I admitting all symmetry classes.

Even for an ideal gas, in which the interaction energy between molecules can be neglected, deviations from the ordinary equation of state have to be expected for the reason that only *one* class of symmetry is possible as soon as the mean De Broglie wavelength of a gas molecule becomes of an order of magnitude comparable with the average distance between two molecules, that is, for small temperatures and large densities. For the antisymmetrical class the statistical consequences have been derived by Fermi and Dirac[13], for the symmetrical class the same had been done already before the discovery of the new quantum mechanics by Einstein and Bose[14]. The former case could be applied to the electrons in a metal and could be used for the interpretation of magnetic and other properties of metals.

As soon as the symmetry classes for electrons were cleared, the question arose which are the symmetry classes for other particles. One example for particles with symmetrical wave functions only (assumption II) was already known long ago, namely the photons. This is not only an immediate consequence of Planck's derivation of the spectral distribution of the radiation energy in the thermodynamical equilibrium, but it is also necessary for the applicability of the classical field concepts to light waves in the limit where a large and not accurately fixed number of photons is present in a single quantum state. We note that the symmetrical class for photons occurs together with the integer value 1 for their spin, while the antisymmetrical class for the electron occurs together with the half-integer value $\frac{1}{2}$ for the spin.

The important question of the symmetry classes for nuclei, however, had still to be investigated. Of course the symmetry class refers here also to the permutation of both the space coordinates and the spin indices of two like nuclei. The spin index can assume $2I + 1$ values if I is the spin-quantum number of the nucleus which can be either an integer or a half-integer. I may include the historical remark that already in 1924, before the electron spin

was discovered, I proposed to use the assumption of a nuclear spin to interpret the hyperfine-structure of spectral lines[15]. This proposal met on the one hand strong opposition from many sides but influenced on the other hand Goudsmit and Uhlenbeck in their claim of an electron spin. It was only some years later that my attempt to interpret the hyperfine-structure could be definitely confirmed experimentally by investigations in which also Zeeman himself participated and which showed the existence of a magneto-optic transformation of the hyperfine-structure as I had predicted it. Since that time the hyperfine-structure of spectral lines became a general method of determining the nuclear spin.

In order to determine experimentally also the symmetry class of the nuclei, other methods were necessary. The most convenient, although not the only one, consists in the investigation of band spectra due to a molecule with two like atoms[16]. It could easily be derived that in the ground state of the electron configuration of such a molecule the states with even and odd values of the rotational quantum number are symmetric and antisymmetric respectively for a permutation of the space coordinates of the two nuclei. Further there exist among the $(2I+1)^2$ spin states of the pair of nuclei, $(2I+1)(I+1)$ states symmetrical and $(2I+1)I$ states antisymmetrical in the spins, since the $(2I+1)$ states with two spins in the same direction are necessarily symmetrical. Therefore the conclusion was reached: If the total wave function of space coordinates and spin indices of the nuclei is symmetrical, the ratio of the weight of states with an even rotational quantum number to the weight of states with an odd rotational quantum number is given by $(I+1):I$. In the reverse case of an antisymmetrical total wave function of the nuclei, the same ratio is $I:(I+1)$. Transitions between one state with an even and another state with an odd rotational quantum number will be extremely rare as they can only be caused by an interaction between the orbital motions and the spins of the nuclei. Therefore the ratio of the weights of the rotational states with different parity will give rise to two different systems of band spectra with different intensities, the lines of which are alternating.

The first application of this method was the result that the protons have the spin $\frac{1}{2}$ and fulfill the exclusion principle just as the electrons. The initial difficulties to understand quantitatively the specific heat of hydrogen molecules at low temperatures were removed by Dennison's hypothesis[17], that at this low temperature the thermal equilibrium between the two modifications of the hydrogen molecule (ortho-H_2: odd rotational quantum numbers,

parallel proton spins; para-H_2: even rotational quantum numbers, antiparallel spins) was not yet reached. As you know, this hypothesis was later, confirmed by the experiments of Bonhoeffer and Harteck and of Eucken, which showed the theoretically predicted slow transformation of one modification into the other.

Among the symmetry classes for other nuclei those with a different parity of their mass number M and their charge number Z are of a particular interest. If we consider a compound system consisting of numbers A_1, A_2, ... of different constituents, each of which is fulfilling the exclusion principle, and a number S of constituents with symmetrical states, one has to expect symmetrical or antisymmetrical states if the sum $A_1 + A_2 + \ldots$ is even or odd. This holds regardless of the parity of S. Earlier one tried the assumption that nuclei consist of protons and electrons, so that M is the number of protons, $M - Z$ the number of electrons in the nucleus. It had to be expected then that the parity of Z determines the symmetry class of the whole nucleus. Already for some time the counter-example of nitrogen has been known to have the spin 1 and symmetrical states[18]. After the discovery of the neutron, the nuclei have been considered, however, as composed of protons and neutrons in such a way that a nucleus with mass number M and charge number Z should consist of Z protons and $M - Z$ neutrons. In case the neutrons would have symmetrical states, one should again expect that the parity of the charge number Z determines the symmetry class of the nuclei. If, however, the neutrons fulfill the exclusion principle, it has to be expected that the parity of M determines the symmetry class: For an even M, one should always have symmetrical states, for an odd M, antisymmetrical ones. It was the latter rule that was confirmed by experiment without exception, thus proving that the neutrons fulfill the exclusion principle.

The most important and most simple crucial example for a nucleus with a different parity of M and Z is the heavy hydrogen or deuteron with $M = 2$ and $Z = 1$ which has symmetrical states and the spin $I = 1$, as could be proved by the investigation of the band spectra of a molecule with two deuterons[19]. From the spin value 1 of the deuteron can be concluded that the neutron must have a half-integer spin. The simplest possible assumption that this spin of the neutron is equal to $\frac{1}{2}$, just as the spin of the proton and of the electron, turned out to be correct.

There is hope, that further experiments with light nuclei, especially with protons, neutrons, and deuterons will give us further information about the nature of the forces between the constituents of the nuclei, which, at present,

is not yet sufficiently clear. Already now we can say, however, that these interactions are fundamentally different from electromagnetic interactions. The comparison between neutron–proton scattering and proton–proton scattering even showed that the forces between these particles are in good approximation the same, that means independent of their electric charge. If one had only to take into account the magnitude of the interaction energy, one should therefore expect a stable di-proton or ${}_2^2\mathrm{He}$ ($M = 2, Z = 2$) with nearly the same binding energy as the deuteron. Such a state is, however, forbidden by the exclusion principle in accordance with experience, because this state would acquire a wave function symmetric with respect to the two protons. This is only the simplest example of the application of the exclusion principle to the structure of compound nuclei, for the understanding of which this principle is indispensable, because the constituents of these heavier nuclei, the protons and the neutrons, fullfil it.

In order to prepare for the discussion of more fundamental questions, we want to stress here a law of Nature which is generally valid, namely, the connection between spin and symmetry class. *A half-integer value of the spin quantum number is always connected with antisymmetrical states (exclusion principle), an integer spin with symmetrical states.* This law holds not only for protons and neutrons but also for protons and electrons. Moreover, it can easily be seen that it holds for compound systems, if it holds for all of its constituents. If we search for a theoretical explanation of this law, we must pass to the discussion of relativistic wave mechanics, since we saw that it can certainly not be explained by non-relativistic wave mechanics.

We first consider classical fields[20], which, like scalars, vectors, and tensors transform with respect to rotations in the ordinary space according to a one-valued representation of the rotation group. We may, in the following, call such fields briefly «one-valued» fields. So long as interactions of different kinds of field are not taken into account, we can assume that all field components will satisfy a second-order wave equation, permitting a superposition of plane waves as a general solution. Frequency and wave number of these plane waves are connected by a law which, in accordance with De Broglie's fundamental assumption, can be obtained from the relation between energy and momentum of a particle claimed in relativistic mechanics by division with the constant factor $\hbar$ equal to Planck's constant divided by 2π. Therefore, there will appear in the classical field equations, in general, a new constant μ with the dimension of a reciprocal length, with which the

rest-mass m in the particle picture is connected by $m = \hbar\,\mu/c$, where c is the vacuum-velocity of light. From the assumed property of one-valuedness of the field it can be concluded, that the number of possible plane waves for a given frequency, wave number and direction of propagation, is for a non-vanishing μ always odd. Without going into details of the general definition of spin, we can consider this property of the polarization of plane waves as characteristic for fields which, as a result of their quantization, give rise to integer spin values.

The simplest cases of one-valued fields are the scalar field and a field consisting of a four-vector and an antisymmetric tensor like the potentials and field strengths in Maxwell's theory. While the scalar field is simply fulfilling the usual wave equation of the second order in which the term proportional to μ^2 has to be included, the other field has to fulfill equations due to Proca which are a generalization of Maxwell's equations which become in the particular case $\mu = 0$. It is satisfactory that for these simplest cases of one-valued fields the energy density is a positive definite quadratic form of the field-quantities and their first derivatives at a certain point. For the general case of one-valued fields it can at least be achieved that the total energy after integration over space is always positive.

The field components can be assumed to be either real or complex. For a complex field, in addition to energy and momentum of the field, a four-vector can be defined which satisfies the continuity equation and can be interpreted as the four-vector of the electric current. Its fourth component determines the electric charge density and can assume both positive and negative values. It is possible that the charged mesons observed in cosmic rays have integral spins and thus can be described by such a complex field. In the particular case of real fields this four-vector of current vanishes identically.

Especially in view of the properties of the radiation in the thermodynamical equilibrium in which specific properties of the field sources do not play any role, it seemed to be justified first to disregard in the formal process of field quantization the interaction of the field with the sources. Dealing with this problem, one tried indeed to apply the same mathematical method of passing from a classical system to a corresponding system governed by the laws of quantum mechanics which has been so successful in passing from classical point mechanics to wave mechanics. It should not be forgotten, however, that a field can only be observed with help of its interaction with test bodies which are themselves again sources of the field.

The result of the formal process of field quantization were partly very

encouraging. The quantized wave fields can be characterized by a wave function which depends on an infinite sequence of (non-negative) integers as variables. As the total energy and the total momentum of the field and, in case of complex fields, also its total electric charge turn out to be linear functions of these numbers, they can be interpreted as the number of particles present in a specified state of a single particle. By using a sequence of configuration spaces with a different number of dimensions corresponding to the different possible values of the total number of particles present, it could easily be shown that this description of our system by a wave function depending on integers is equivalent to an ensemble of particles with wave functions symmetrical in their configuration spaces.

Moreover Bohr and Rosenfeld[21] proved in the case of the electromagnetic field that the uncertainty relations which result for the average values of the field strengths over finite space-time regions from the formal commutation rules of this theory have a direct physical meaning so long as the sources can be treated classically and their atomistic structure can be disregarded. We emphasize the following property of these commutation rules: All physical quantities in two world points, for which the four-vector of their joining straight line is spacelike commute with each other. This is indeed necessary for physical reasons because any disturbance by measurements in a world point P_1 can only reach such points P_2 for which the vector P_1P_2 is timelike, that is, for which $c(t_1 - t_2) > r_{12}$. The points P_2 with a spacelike vector P_1P_2 for which $c(t_1 - t_2) < r_{12}$ cannot be reached by this disturbance and measurements in P_1 and P_2 can then never influence each other.

This consequence made it possible to investigate the logical possibility of particles with integer spin which would obey the exclusion principle. Such particles could be described by a sequence of configuration spaces with different dimensions and wave functions antisymmetrical in the coordinates of these spaces or also by a wave function depending on integers again to be interpreted as the number of particles present in specified states which now can only assume the values 0 or 1. Wigner and Jordan[22] proved that also in this case operators can be defined which are functions of the ordinary space-time coordinates and which can be applied to such a wave function. These operators do not fulfil any longer commutation rules: instead of the difference, the *sum* of the two possible products of two operators, which are distinguished by the different order of its factors, is now fixed by the mathematical conditions the operators have to satisfy. The simple change of the sign in these conditions changes entirely the physical meaning of the for-

malism. In the case of the exclusion principle there can never exist a limiting case where such operators can be replaced by a classical field. Using this formalism of Wigner and Jordan I could prove under very general assumptions that a relativistic invariant theory describing systems of like particles with integer spin obeying the exclusion principle would always lead to the non-commutability of physical quantities joined by a spacelike vector[23]. This would violate a reasonable physical principle which holds good for particles with symmetrical states. In this way, by combination of the claims of relativistic invariance and the properties of field quantization, one step in the direction of an understanding of the connection of spin and symmetry class could be made.

The quantization of one-valued complex fields with a non-vanishing four-vector of the electric current gives the further result that particles both with positive and negative electric charge should exist and that they can be annihilated and generated in external electromagnetic fields[22]. This pair-generation and annihilation claimed by the theory makes it necessary to distinguish clearly the concept of charge density and of particle density. The latter concept does not occur in a relativistic wave theory either for fields carrying an electric charge or for neutral fields. This is satisfactory since the use of the particle picture and the uncertainty relations (for instance by analyzing imaginative experiments of the type of the γ-ray microscope) gives also the result that a localization of the particle is only possible with limited accuracy[24]. This holds both for the particles with integer and with half-integer spins. In a state with a mean value E of its energy, described by a wave packet with a mean frequency $\nu = E/h$, a particle can only be localized with an error $\Delta x > hc/E$ or $\Delta x > c/\nu$. For photons, it follows that the limit for the localization is the wavelength; for a particle with a finite rest-mass m and a characteristic length $\mu^{-1} = \hbar/mc$, this limit is in the rest system of the center of the wave packet that describes the state of the particles given by $\Delta x > \hbar/mc$ or $\Delta x > \mu^{-1}$.

Until now I have mentioned only those results of the application of quantum mechanics to classical fields which are satisfactory. We saw that the statements of this theory about averages of field strength over finite space-time regions have a direct meaning while this is not so for the values of the field strength at a certain point. Unfortunately in the classical expression of the energy of the field there enter averages of the squares of the field strengths over such regions which cannot be expressed by the averages of the field strengths themselves. This has the consequence that the zero-point energy

of the vacuum derived from the quantized field becomes infinite, a result which is directly connected with the fact that the system considered has an infinite number of degrees of freedom. It is clear that this zero-point energy has no physical reality, for instance it is not the source of a gravitational field. Formally it is easy to subtract constant infinite terms which are independent of the state considered and never change; nevertheless it seems to me that already this result is an indication that a fundamental change in the concepts underlying the present theory of quantized fields will be necessary.

In order to clarify certain aspects of relativistic quantum theory I have discussed here, different from the historical order of events, the one-valued fields first. Already earlier Dirac[25] had formulated his relativistic wave equations corresponding to material particles with spin $\frac{1}{2}$ using a pair of so-called spinors with two components each. He applied these equations to the problem of one electron in an electromagnetic field. In spite of the great success of this theory in the quantitative explanation of the fine structure of the energy levels of the hydrogen atom and in the computation of the scattering cross section of one photon by a free electron, there was one consequence of this theory which was obviously in contradiction with experience. The energy of the electron can have, according to the theory, both positive and negative values, and, in external electromagnetic fields, transitions should occur from states with one sign of energy to states with the other sign. On the other hand there exists in this theory a four-vector satisfying the continuity equation with a fourth component corresponding to a density which is definitely positive.

It can be shown that there is a similar situation for all fields, which, like the spinors, transform for rotations in ordinary space according to two-valued representations, thus changing their sign for a full rotation. We shall call briefly such quantities « two-valued ». From the relativistic wave equations of such quantities one can always derive a four-vector bilinear in the field components which satisfies the continuity equation and for which the fourth component, at least after integration over the space, gives an essentially positive quantity. On the other hand, the expression for the total energy can have both the positive and the negative sign.

Is there any means to shift the minus sign from the energy back to the density of the four-vector? Then the latter could again be interpreted as charge density in contrast to particle density and the energy would become positive as it ought to be. You know that Dirac's answer was that this could actually be achieved by application of the exclusion principle. In his lecture

delivered here in Stockholm[10] he himself explained his proposal of a new interpretation of his theory, according to which in the actual vacuum all the states of negative energy should be occupied and only deviations of this state of smallest energy, namely holes in the sea of these occupied states are assumed to be observable. It is the exclusion principle which guarantees the stability of the vacuum, in which all states of negative energy are occupied. Furthermore the holes have all properties of particles with positive energy and positive electric charge, which in external electromagnetic fields can be produced and annihilated in pairs. These predicted positrons, the exact mirror images of the electrons, have been actually discovered experimentally.

The new interpretation of the theory obviously abandons in principle the standpoint of the one-body problem and considers a many-body problem from the beginning. It cannot any longer be claimed that Dirac's relativistic wave equations are the only possible ones but if one wants to have relativistic field equations corresponding to particles, for which the value $\frac{1}{2}$ of their spin is known, one has certainly to assume the Dirac equations. Although it is logically possible to quantize these equations like classical fields, which would give symmetrical states of a system consisting of many such particles, this would be in contradiction with the postulate that the energy of the system has actually to be positive. This postulate is fulfilled on the other hand if we apply the exclusion principle and Dirac's interpretation of the vacuum and the holes, which at the same time substitutes the physical concept of charge density with values of both signs for the mathematical fiction of a positive particle density. A similar conclusion holds for all relativsitic wave equations with two-valued quantities as field components. This is the other step (historically the earlier one) in the direction of an understanding of the connection between spin and symmetry class.

I can only shortly note that Dirac's new interpretation of empty and occupied states of negative energy can be formulated very elegantly with the help of the formalism of Jordan and Wigner mentioned before. The transition from the old to the new interpretation of the theory can indeed be carried through simply by interchanging the meaning of one of the operators with that of its hermitian conjugate if they are applied to states originally of negative energy. The infinite « zero charge » of the occupied states of negative energy is then formally analogous to the infinite zero-point energy of the quantized one-valued fields. The former has no physicial reality either and is not the source of an electromagnetic field.

In spite of the formal analogy between the quantization of the one-valued fields leading to ensembles of like particles with symmetrical states and to particles fulfilling the exclusion principle described by two-valued operator quantities, depending on space and time coordinates, there is of course the fundamental difference that for the latter there is no limiting case, where the mathematical operators can be treated like classical fields. On the other hand we can expect that the possibilities and the limitations for the applications of the concepts of space and time, which find their expression in the different concepts of charge density and particle density, will be the same for charged particles with integer and with half-integer spins.

The difficulties of the present theory become much worse, if the interaction of the electromagnetic field with matter is taken into consideration, since the well-known infinities regarding the energy of an electron in its own field, the so-called self-energy, then occur as a result of the application of the usual perturbation formalism to this problem. The root of this difficulty seems to be the circumstance that the formalism of field quantization has only a direct meaning so long as the sources of the field can be treated as continuously distributed, obeying the laws of classical physics, and so long as only averages of field quantities over finite space-time regions are used. The electrons themselves, however, are essentially non-classical field sources.

At the end of this lecture I may express my critical opinion, that a correct theory should neither lead to infinite zero-point energies nor to infinite zero charges, that it should not use mathematical tricks to subtract infinities or singularities, nor should it invent a «hypothetical world» which is only a mathematical fiction before it is able to formulate the correct interpretation of the actual world of physics.

From the point of view of logic, my report on «Exclusion principle and quantum mechanics» has no conclusion. I believe that it will only be possible to write the conclusion if a theory will be established which will determine the value of the fine-structure constant and will thus explain the atomistic structure of electricity, which is such an essential quality of all atomic sources of electric fields actually occurring in Nature.

1. A. Landé, *Z. Physik*, 5 (1921) 231 and *Z. Physik*, 7 (1921) 398, *Physik. Z.*, 22 (1921) 417.
2. W. Pauli, *Z. Physik*, 16 (1923) 155.
3. W. Pauli, *Z. Physik*, 31 (1925) 373.
4. E. C. Stoner, *Phil. Mag.*, 48 (1924) 719.
5. W. Pauli, *Z. Physik*, 31 (1925) 765.
6. S. Goudsmit and G. Uhlenbeck, *Naturwiss.*, 13 (1925) 953, *Nature*, 117 (1926) 264.
7. L. H. Thomas, *Nature*, 117 (1926) 514, and *Phil. Mag.*, 3 (1927) 1. Compare also J. Frenkel, *Z. Physik*, 37 (1926) 243.
8. Compare *Rapport du Sixième Conseil Solvay de Physique, Paris, 1932*, pp. 217–225.
9. For this earlier stage of the history of the exclusion principle compare also the author's note in *Science*, 103 (1946) 213, which partly coincides with the first part of the present lecture.
10. The Nobel Lectures of W. Heisenberg, E. Schrödinger, and P. A. M. Dirac are collected in *Die moderne Atomtheorie*, Leipzig, 1934.
11. The articles of N. Bohr are collected in *Atomic Theory and the Description of Nature*, Cambridge University Press, 1934. See also his article «Light and Life», *Nature*, 131 (1933) 421, 457.
12. W. Heisenberg, *Z. Physik*, 38 (1926) 411 and 39 (1926) 499.
13. E. Fermi, *Z. Physik*, 36 (1926) 902.
P. A. M. Dirac, *Proc. Roy. Soc. London*, A 112 (1926) 661.
14. S. N. Bose, *Z. Physik*, 26 (1924) 178 and 27 (1924) 384.
A. Einstein, *Berl. Ber.*, (1924) 261; (1925) 1, 18.
15. W. Pauli, *Naturwiss.*, 12 (1924) 741.
16. W. Heisenberg, *Z. Physik*, 41 (1927) 239, F. Hund, *Z. Physik*, 42 (1927) 39.
17. D. M. Dennison, *Proc. Roy. Soc. London*, A 115 (1927) 483.
18. R. de L. Kronig, *Naturwiss.*, 16 (1928) 335.
W. Heitler und G. Herzberg, *Naturwiss.*, 17 (1929) 673.
19. G. N. Lewis and M. F. Ashley, *Phys. Rev.*, 43 (1933) 837.
G. M. Murphy and H. Johnston, *Phys. Rev.*, 45 (1934) 550 and 46 (1934) 95.
20. Compare for the following the author's report in *Rev. Mod. Phys.*, 13 (1941) 203, in which older literature is given. See also W. Pauli and V. Weisskopf, *Helv. Phys. Acta*, 7 (1934) 809.
21. N. Bohr and L. Rosenfeld, *Kgl. Danske Videnskab. Selskab. Mat. Fys. Medd.*, 12 [8](1933).
22. P. Jordan and E. Wigner, *Z. Physik*, 47 (1928) 631.
Compare also V. Fock, *Z. Physik*, 75 (1932) 622.
23. W. Pauli, *Ann. Inst. Poincaré*, 6 (1936) 137 and *Phys. Rev.*, 58 (1940) 716.
24. L. Landau and R. Peierls, *Z. Physik*, 69 (1931) 56.
Compare also the author's article in *Handbuch der Physik*, 24, Part 1, 1933, Chap. A, § 2.
25. P. A. M. Dirac, *Proc. Roy. Soc. London*, A 117 (1928) 610.

Biography

Wolfgang Pauli was born on April 25th, 1900 in Vienna. He received his early education in Vienna before studying at the University of Munich under Arnold Sommerfeld. He obtained his doctor's degree in 1921 and spent a year at the University of Göttingen as assistant to Max Born and a further year with Niels Bohr at Copenhagen. The years 1923–1928 were spent as a lecturer at the University of Hamburg before his appointment as Professor of Theoretical Physics at the Federal Institute of Technology in Zurich. During 1935–1936, he was visiting Professor at the Institute for Advanced Study, Princeton, New Jersey and he had similar appointments at the University of Michigan (1931 and 1941) and Purdue University (1942). He was elected to the Chair of Theoretical Physics at Princeton in 1940 but he returned to Zurich at the end of World War II.

Pauli was outstanding among the brilliant mid-twentieth century school of physicists. He was recognized as one of the leaders when, barely out of his teens and still a student, he published a masterly exposition of the theory of relativity. His exclusion principle, which is often quoted bearing his name, crystallized the existing knowledge of atomic structure at the time it was postulated and it led to the recognition of the two-valued variable required to characterize the state of an electron. Pauli was the first to recognize the existence of the neutrino, an uncharged and massless particle which carries off energy in radioactive β-disintegration; this came at the beginning of a great decade, prior to World War II, for his centre of research in theoretical physics at Zurich.

Pauli helped to lay the foundations of the quantum theory of fields and he participated actively in the great advances made in this domain around 1945. Earlier, he had further consolidated field theory by giving proof of the relationship between spin and « statistics » of elementary particles. He has written many articles on problems of theoretical physics, mostly quantum mechanics, in scientific journals of many countries; his *Theory of Relativity* appears in the *Enzyklopaedie der Mathematischen Wissenschaften*, Volume 5, Part 2 (1920), his *Quantum Theory* in *Handbuch der Physik*, Vol. 23 (1926),

and his *Principles of Wave Mechanics* in *Handbuch der Physik*, Vol. 24 (1933).

Pauli was a Foreign Member of the Royal Society of London and a member of the Swiss Physical Society, the American Physical Society and the American Association for the Advancement of Science. He was awarded the Lorentz Medal in 1930.

Wolfgang Pauli married Franciska Bertram on April 4th, 1934. He died in Zurich on December 15th, 1958.

Physics 1946

PERCY WILLIAMS BRIDGMAN

«for the invention of an apparatus to produce extremely high pressures, and for the discoveries he made therewith in the field of high-pressure physics»

Physics 1946

Presentation Speech by Professor A. E. Lindh, member of the Nobel Committee for Physics

Your Majesty, Your Royal Highnesses, Ladies and Gentlemen.

The earliest known attempts to attain high pressures and to study various properties of matter under the influence of these pressures date from the beginning and middle of the 17th century. The experiments, which were carried out by extremely primitive methods, aimed in the first place at throwing light on the compressibility of liquids. These investigations did not become of a more scientific and systematic nature until the beginning of the last century, although at first they had the same limited aims as before. However, they were gradually extended to other fields. As an example may be mentioned attempts to compress a number of gases at high pressure, when great divergencies from Boyle's law revealed themselves, and further, investigations of the effect of pressure on the refractivity of water, the effect of pressure on the resistance of an electric conductor. Most important of all was in 1861 Andrews' discovery of the critical phenomena in gases.

A period of intensive research began after this last discovery, a period which lasted until the beginning of the 1890's. During this period the leading research workers in this field were the two French physicists, Cailletet and Amagat, of whom the former was active during an earlier phase of the period mentioned. These two scientists, especially the latter, made important contributions towards improving the technique for attaining high pressures and worked out reliable methods for measuring them. Amagat developed a special technique for ensuring effective sealings or packings, a fundamental problem when working on high pressures. Thanks to this technique, Amagat succeeded in obtaining constant pressures of 3,000 kg/cm^2 and more. His contributions were of the greatest importance for further work, and subsequently a large number of research workers in different countries devoted themselves to the study of high pressures. Although very comprehensive work in this field was done, no considerable progress was made in the matter of an improved technique. The limit remained at 3,000 kg/cm^2, and it was not until after 1905 that it began to make gigantic leaps upwards. The credit for this is due in the first place to Professor P. W. Bridgman, who is today

to receive the 1946 Nobel Prize for Physics for his invention of an apparatus for obtaining extremely high pressures, and for the discoveries he has made with it within that field of physics.

According to Professor Bridgman himself, it was by a mere chance that he came to devote his activities to high pressures. At about the year 1905, he began to study certain optical phenomena under the influence of pressure. During his experiments the apparatus, parts of which were made of glass, exploded, and an essential part of it was destroyed and had to be replaced. In the interval, Bridgman tried to find another use for the actual pressure apparatus, and while working out a sealing device for the pressure chamber, he found that the sealing device he had constructed functioned far better than he had at first imagined, for the efficiency of the sealing proved to increase as the pressure increased, and there was no perceptible leakage. A new pressure range had presented itself, a range which was not, as in Amagat's experiments, limited by leakage, but only by the strength of the material of which the pressure apparatus was made. After the problem of leakage had been solved, an advance towards higher pressures came to be essentially a question of materials.

Even in his earliest experiments Bridgman succeeded in arriving at pressures of 20,000 kg/cm^2. In the earlier investigations, however, the pressure was kept at a lower value, in order to avoid deformation of the material then used. Considerable time was devoted both to investigating the material and to different methods of making accurate determinations of the pressures. From his first successful attempts to pass Amagat's pressure limit of 3,000 kg/cm^2, Bridgman has step by step, by means of his brilliant apparatus and skilful use of the resources of modern technics, extended the pressure range, and has made pressures up to 100,000 kg/cm^2 available for research work. In certain cases pressures of between 400,000 and 500,000 kg/cm^2 have been attained.

The essential features of the Bridgman pressure apparatus are two containers, communicating with each other by means of strong connecting channels. The whole system is filled with an appropriate fluid. In one of the containers (the pressure chamber itself) the pressure fluid is subjected, by means of a movable piston, to a great pressure, which is transmitted by the fluid to the other container, the actual experimental chamber. This last-mentioned part varies in accordance with the nature of the projected investigations.

Working on the principle that the resistance of a vessel which is subjected

to internal pressure is increased if at the same time it is subjected to external pressure, Bridgman used double high-pressure vessels for ranges from ca. 30,000 to 100,000 kg/cm². The internal pressure vessel with an external conical surface is fitted into a corresponding bore in a reinforcement cylinder and equipped with a cylindrical channel, where the material which is to be subjected to pressure is placed between two pistons working from opposite ends. As material for the high-pressure vessel in the range from 50,000 to 100,000 kg/cm², Bridgman used the extremely hard tungsten carbide, or, as it is called, carboloy, which is subject to the least possible deformation.

The investigations with the highest pressures, between 400,000 and 500,000 kg/cm², have been made with the help of carboloy pistons, the effective surface being 3 mm in diameter. This small size is a necessary result of the expansion of the pressure area under these high pressures, and therefore the amount of matter which can be compressed is extremely small. With a pressure of 425,000 kg/cm² Bridgman obtained compressed material in the form of small thin flakes. An X-ray analysis of these showed that, under the influence of the pressure, the structure had passed from crystalline to non-crystalline form, i.e. the substance had become amorphous.

Bridgman's research work has not been directed only towards attaining record high pressures. The gradual advance towards higher and ever higher pressures was immediately taken advantage of for investigations in fields where the impossibility of attaining higher pressures had previously put a stop to research work. Bridgman's important discoveries in the course of this work are so many that here it is only possible to touch upon them briefly.

The majority of Bridgman's earlier investigations were carried out within the range up to 12,000 kg/cm². The first more comprehensive investigations had to do with the solid and fluid conditions, and these investigations were subsequently extended to ranges up to 50,000 kg/cm², new modifications of different substances being discovered, *inter alia*, of both ordinary and heavy water in solid form, and altogether seven different modifications of ice are known. Further, two new modifications of phosphorus have been discovered, one stable form – the so-called black phosphorus – and one unstable form. By means of investigations of compressibility at pressures right up to 100,000 kg/cm², a large number of polymorphous substances have been discovered. A great deal of work has been devoted to meticulous investigations of the pressure effect on electric resistance, and here, *inter alia*, the existence of a resistance minimum for certain metals at very high pressures has been

established. Bridgman's interest has also been directed towards other spheres. Thus, investigations have been made into the pressure effect on thermo-electric phenomena, on the conduction of heat in gases, on the viscosity of fluids, which have led to discoveries, significant both scientifically and technically. This applies also to his work on the effect of pressure on the elastic properties of solid bodies. These contributions were all made in fields which had previously not attracted much interest. Attention should also be called to the extensive and difficult investigations of materials, which were a necessary precondition for the successful advances towards higher pressures, investigations which are of the greatest importance for further work in the field of high pressures.

Finally, attention should be called to the immense scientific value of the impressive collection of data regarding the properties of matter at high pressures which Bridgman has assembled during his long and important research activities in the field of high pressure physics.

Professor Bridgman. In awarding you this year's Nobel Prize for Physics, The Royal Swedish Academy of Sciences desires to express its unreserved acknowledgement of your outstanding pioneer work in the field of high-pressure physics. By means of your ingenious apparatus, combined with a brilliant experimental technique, you have, by your intense research work and the resulting manifold and remarkable discoveries, very greatly enriched our knowledge of the properties of matter at high pressures.

On behalf of The Royal Swedish Academy of Sciences, I congratulate you on your important and successful work in the service of science, and I now ask you to receive your Nobel Prize from the hands of His Majesty the King.

PERCY W. BRIDGMAN

General survey of certain results in the field of high-pressure physics

Nobel Lecture, December 11, 1946

In this lecture I shall attempt to present a general survey of those parts of the field of high-pressure physics with which I have had direct contact, dealing first with technical matters of producing and measuring high pressure, and secondly with the physical phenomena which occur under high pressure.

With regard to technique, several different ranges of pressure are to be recognized. The first step was to devise a method of packing which should be without leak, since leak had limited the range of previous experiments. A packing was devised, shown in Fig. 1, which automatically becomes tighter the higher the pressure, so that any pressure is accessible up to the strength of the containing vessels. If the vessels are made of one-piece construction,

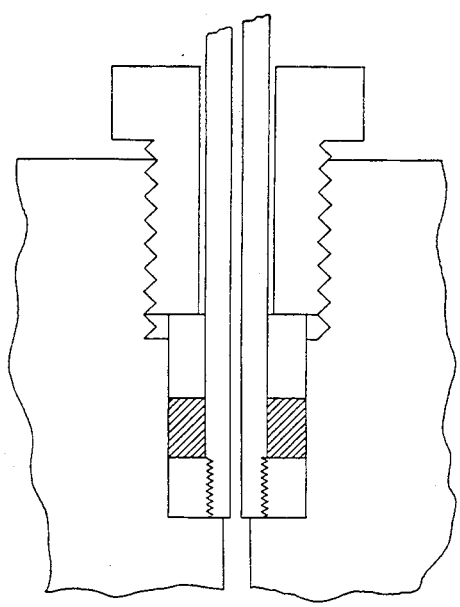


Fig. 1. The general scheme of the packing by which pressure in the soft packing materials is automatically maintained a fixed percentage higher than in the liquid.

from the best heat-treated alloy steels, it is possible to reach pressures of 12,000 kg/cm^2 as a routine matter and on occasion for short intervals of time as high as 20,000. For many years my work was confined to this range, and in this range it proved feasible to measure nearly all the ordinary physical properties of substances. The next step was to give the pressure vessel external support which increases in magnitude at the same time the internal pressure increases. A simple method of doing is to make the external surface of the pressure vessel conical in shape, and to push it into a heavy collar with a force which increases as the internal pressure increases, as illustrated in Fig. 2. With apparatus of this kind it is possible to make routine experiments up to 30,000 kg/cm^2 with volumes of the order of 15 cm^3, to get electrically insulated leads into the apparatus, and practically to repeat all the former work in the range to 12,000. I am still engaged in carrying out this program. An extension of the same technique on a smaller scale with capacities of the order of 0.5 cm^3 can be made up to 50,000 kg/cm^2. In this range all ordinary liquids freeze solid, electrically insulated leads cannot be got into the apparatus, and the phenomena which can be studied are limited to various volume effects, such as compressibilities and phase changes, including fusions and polymorphic transitions.

The external support of the vessel is only one of the factors that makes possible the extension of range from 12,000 to 50,000. No steel piston will support as much as 50,000; carboloy, however, the recently developed substance for tools formed by cementing a fine powder of tungsten carbide with

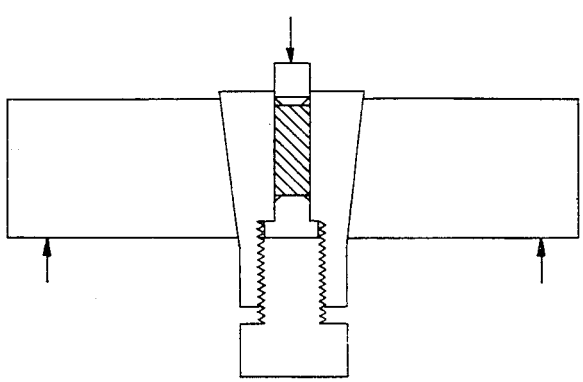


Fig. 2. Illustrating the general principle of the method for giving external support to the pressure vessel in such a way that support increases automatically with the increase of internal pressure.

cobalt, fortunately proves to have a compressive strength high enough for the purpose.

The next step in extension of range, from 50,000 to 100,000 kg/cm², demands still more effective support of the pressure vessel. This is done by immersing the entire pressure vessel in a fluid under pressures ranging up to

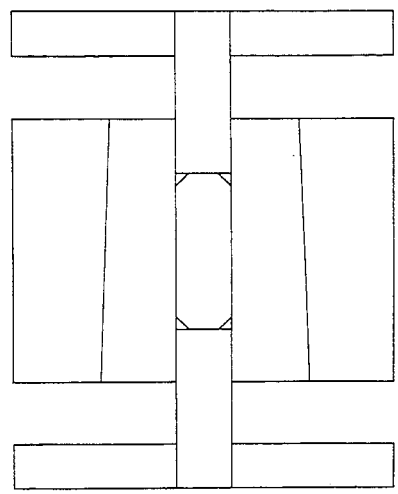


Fig. 3. The miniature apparatus for reaching 100,000 kg/cm².

30,000 kg/cm². The pressure apparatus has to be made still smaller; the pistons are only 1.6 mm in diameter, and the capacity is only a few cubic millimeters. The pressure cylinder itself, as well as the pistons, is now made of carboloy with an external jacket of shrunk-on steel to give it greater strength. The piezometer is illustrated in Fig. 3. Even with this type of construction so great an extension of range as from 50,000 to 100,000 would not have been possible if it were not for a fortunate change in the properties of metals under pressure. At pressures of 25,000 kg/cm² ordinary grades of steel become capable of almost indefinite deformation without fracture, so greatly has their ductility been increased, as shown in Fig. 4. Even carboloy loses its normal brittleness and becomes capable of supporting without fracture higher tensile stresses than steel.

Up to the present, the compressibilities and polymorphic transitions of some 30 elements and simple compounds have been studied in the range to 100,000.

Much higher pressures than 100,000 can be reached in very small regions

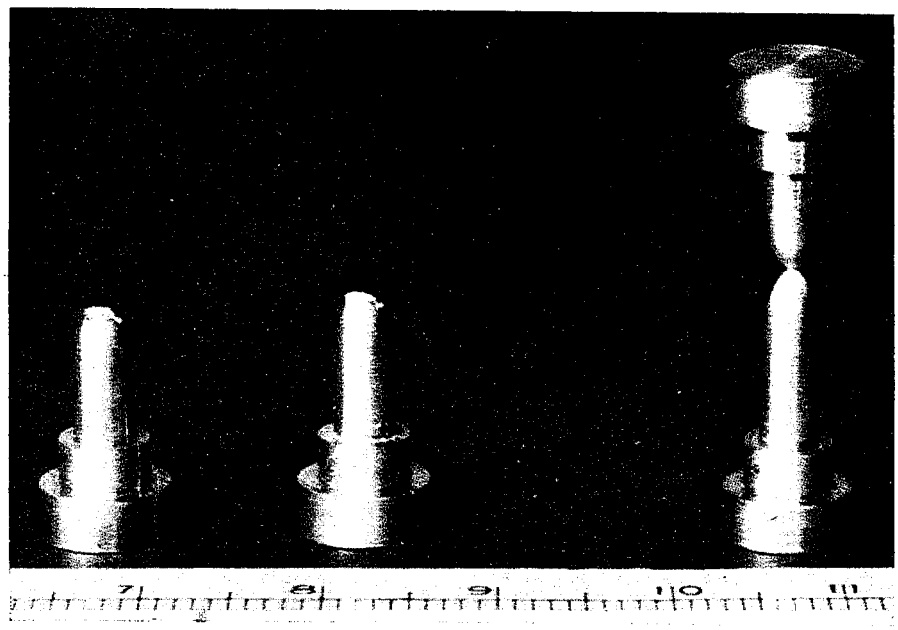


Fig. 4. Illustrating the effect of pressure in increasing the ductility of steel. *On the left*, a piece of mild steel broken in tension at atmospheric pressure. *On the right*, the same steel pulled to a much greater reduction of area without fracture in a liquid at 25,000 kg/cm^2.

by constructing the apparatus entirely of carboloy, but up to the present no particularly important physical results have been attained in this range.

In addition to the problem of attaining the pressures, there is the problem of measuring them and measuring the effects which they produce. This demands in the first place the establishment of various fixed points. In the range up to 30,000 a sufficient number of such points has been established to permit measurements to an accuracy of about 0.1 per cent. A transition of bismuth in the neighborhood of 25,000 gives one such convenient point. An essential part of the measuring technique is the utilization of the change of resistance of manganin under pressure, first suggested by Lisell at Uppsala. Above 30,000 the territory is not so well marked out; it is probable that the measurements to 100,000 have an accuracy of about 2 per cent.

It is natural to think of volume compression as the simplest and most fundamental of all the effects of hydrostatic pressure, and for that reason it will be discussed first here. It is not, however, the simplest to measure experimentally, because the measurements immediately obtained are relative to

the containing vessel, which is itself distorted. Elaborate procedures may be necessary to eliminate the effect of such distortion.

The compression of gases is outside the range of this work; at pressures of 1,000 kg/cm² or more the densities of gases become of the same order of magnitude as those of their liquid phase, and there ceases to be any essential difference between gas and liquid. If the volume of any ordinary liquid is plotted as a function of pressure at constant temperature, a curve will be obtained which at low pressures has a high degree of curvature and a steep tangent, meaning a high compressibility, but as pressure increases the curvature rapidly becomes less and the curve flattens off. In Fig. 5 the volume of a typical liquid, ether, is shown as a function of pressure. For comparison, the curve of the most compressible solid, caesium, is also shown. Two different physical mechanisms are primarily responsible for the different behavior in the low and high pressure ranges. The low range of high compressibility is the range in which the chief effect of pressure is to push the mol-

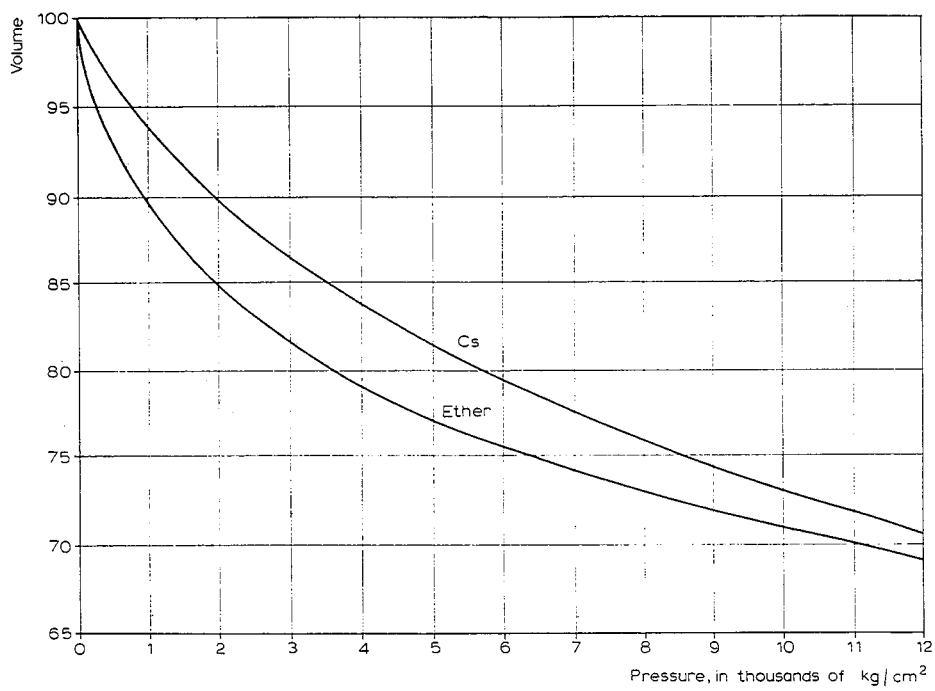


Fig. 5. Volume as a function of pressure for a typical liquid, ether. The corresponding curve is also shown for caesium, the most compressible solid. The liquid is initially much more compressible than the solid, but at higher pressures is less compressible.

ecules into closer contact, eliminating the free spaces between them. In this range individual substances may show large and characteristic individual differences. In the higher range the molecules have been pushed into effective contact, and the compressibility now arises from the decrease of volume of the molecules themselves. This effect persists with comparatively little decrease over a wide range of pressure. This effect is of course present also in the lower range of pressure, but there it is masked by the much larger effect arising from squeezing out the free spaces between the molecules. If one attempts to set up a formula for the effect of pressure on volume on the basis of measurements in the low range only, one will be likely to neglect too much the contribution from the compressibility of the molecules, with the result that the actual volumes at high pressures will be found to be materially smaller than the volumes which would be extrapolated from the low-pressure formulas. This, as a matter of fact, has been a property of practically all the formulas that have been derived from low-pressure data.

At high pressures, the volumes of ordinary organic liquids become surprisingly alike in spite of initial differences. To illustrate the rapid falling off of compressibility with pressure, the volume change in the first 5,000 kg/cm^2 is roughly the same on the average as the volume change between 5,000 and 50,000; the effect is accentuated by the fact that the volume decrement in the latter range often includes the volume discontinuity on freezing.

In the low-pressure range, in which the molecules are being pushed into effective contact, one might expect effects depending on the shapes of the molecules, and that these effects would be highly specific with the liquid. This is the case. In the low-pressure range a great variety of small-scale abnormalities are superposed on the larger-scale uniformities, and these small-scale effects vary greatly from liquid to liquid. Thus there may be sub-ranges of an extent of a few thousand kg/cm^2 in which the compressibility increases with increasing pressure instead of decreasing as is normal, or the thermal expansion may also increase with increasing pressure instead of decreasing. Any satisfactory theory of liquids must ultimately give an account of these small-scale effects but for the present the large-scale effects must have first attention. When the theory of liquids does come to be written, the first step may well be to set up an idealized « perfect liquid » in analogy to the perfect gas which has played so important a role in the theory of gases. The experimental results at high pressures show sufficient uniformity in the behavior of all ordinary organic liquids to indicate that such an idealized perfect liquid is not too far from the actuality.

The compressibility of solids varies over a much wider range than does that of the ordinary liquids; caesium, for example, is 350 times more compressible than diamond. The highest compressibilities among solids, judging from indirect evidence, are probably to be found in solid hydrogen and helium. As in the case of liquids, the compressibility of solids normally drops off with increasing pressure. This would be expected in general because of the operation of a law of « diminishing returns », and is obviously necessary when pressure is raised indefinitely, because if volume continued to decrease at its initial rate it would eventually become negative. For instance, the volume of caesium would become negative at a pressure of only 14,000 kg/cm² if it continued to decrease with pressure at the initial rate. In spite of the fact

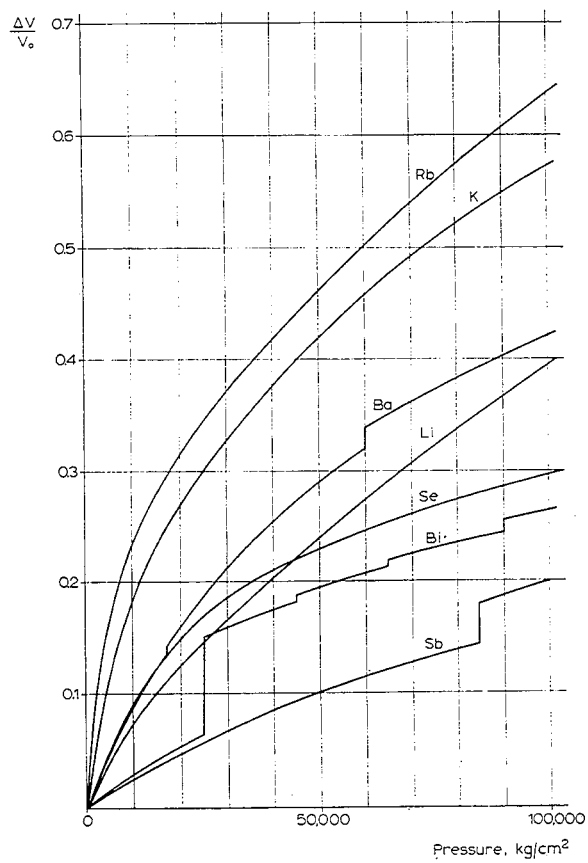


Fig. 6. The volume compression of several elements up to 100,000 kg/cm². The breaks in some of the curves indicate polymorphic transitions.

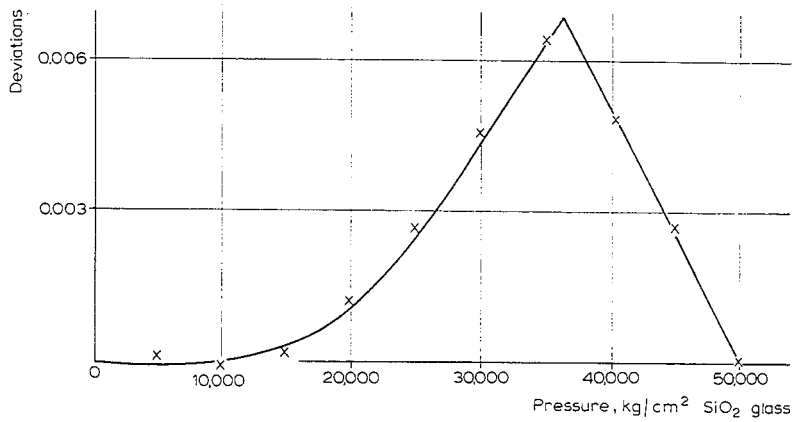


Fig. 7. The deviations from linearity of the volume decrements of quartz glass for pressure increments of 5,000 kg/cm² plotted against pressure. The cusp in the curve marks the change from abnormal to normal behavior.

that the compressibility of solids on the average must decrease with increasing pressure, there is a very marked qualitative difference as compared with liquids. The initial phase of very rapid decrease is absent, and the decrease is spread more uniformly over the entire pressure range. The difference is to be accounted for by the lattice structure of solids; with increasing pressure the atoms retain their position in the lattice with the result that a smaller part of the free space between the atoms is available for occupancy as the centers of the atoms are forced closer together.

The volume decrements of a number of the more compressible solids are shown as a function of pressure up to 100,000 kg/cm² in Fig. 6. The curvature is in general very marked.

There is no thermodynamic necessity that the compressibility should decrease with increasing pressure, although this opinion has sometimes been expressed. Solid substances are known in which the compressibility may increase with increasing pressure over a comparatively wide range of pressure. The most striking example is quartz glass. The compressibility not only increases with pressure, but increases at an accelerating rate. This continues up to 35,000 kg/cm² and then abruptly stops. At this pressure there is a discontinuity in the derivative, a transition of the « second kind » in the nomenclature of Ehrenfest, and from here on compressibility decreases with rising pressure as is normal. The mechanism which is responsible for the low-pressure effect abruptly ceases to act. Fig. 7 shows the relations.

So far we have been considering the effect of pressure on the volume of isotropic substances; this includes substances like glass and all cubic crystals. If the material crystallizes in some non-cubic system, the effects are more complicated. The compressibility is not the same in all directions, so that the shape of bodies composed of such crystals may change under pressure. The differences of compressibility in different directions may be large; thus zinc is eight times as compressible in the direction of the hexagonal axis as at right angles to it. Some difference in this direction might be expected, because the atomic spacing is greater along the axis than at right angles, but no simple consideration would lead to the expectation of differences as large as this. There is even one substance, tellurium, which has a negative compressibility along the axis. That is, when a single crystal of tellurium is subjected to hydrostatic pressure by a fluid in which it is completely immersed, it expands along the axis.

Considerable success has been achieved in calculating theoretically the effect of pressure on the volume of simple solids. The first success was with simple ionic lattices of the type of NaCl by Max Born, who was able to get acceptable values for the lattice spacing and for the initial compressibility. He was not at first successful, however, in reproducing the change of compressibility with pressure, and even today complete success has not been attained in this regard. More complete results have recently been obtained for the alkali metals by applying the methods of wave mechanics. Bardeen has had surprising success in reproducing the entire volume curve over the experimental pressure range for the alkali metals. The calculations are particularly simple here because there is only one free electron per atom, and it turns out that the major part of the effect arises from the increase of kinetic energy of the free electrons arising from their decrease in effective wavelength when the volume is decreased. Other metals, with more free electrons, are more difficult to compute, but it is anticipated that the difficulties are merely difficulties of the complexity of the calculation.

Theory is apparently not yet in a position to attack with much success the problem of non-cubic crystals.

We now consider the discontinuous volume effects arising from phase changes of various sorts. The simplest of these is the effect of pressure on melting. Historically the study of the effect of pressure on melting was approached with the anticipation that the effects would be found to be similar to the effect of pressure on vaporization, and in particular that there would be critical phenomena, so that above a certain pressure and temperature con-

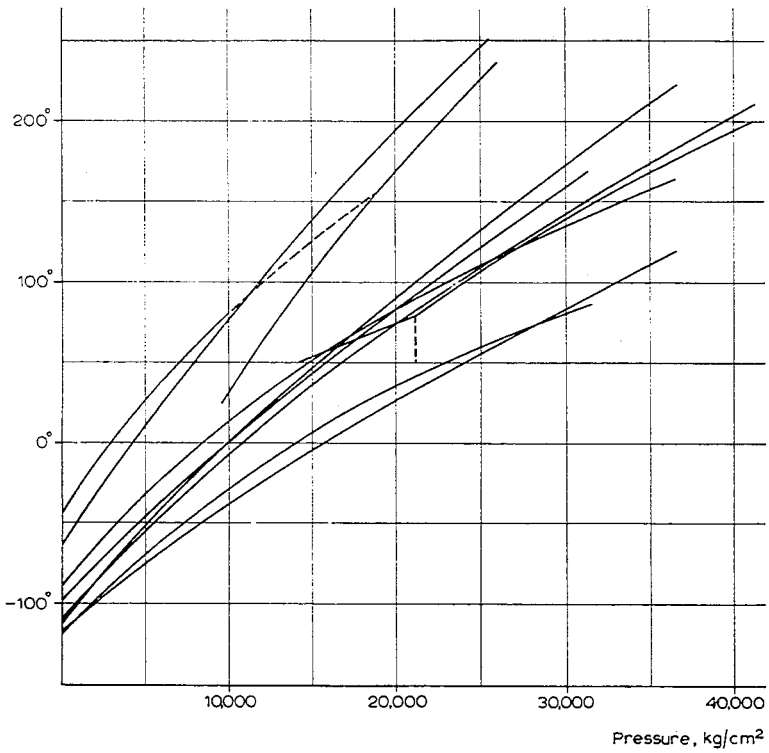


Fig. 8. Melting temperature against pressure for a number of substances. *At 15,000 kg/cm² the order of substances, reading from top down,* is chloroform, chlorobenzene, chlorobenzene (second modification), water (ice VI), 4-butyl alcohol, carbon bisulfide, methylene chloride, *n*-propyl bromide, ethyl bromide, and ethyl alcohol.

tinuous passage would be possible between liquid and solid. It soon appeared, however, that the pressure scale of any such effects must be much more extensive than the scale of the critical effects between liquid and vapor, and that whereas pressures of a few hundred kg/cm² were adequate in the latter case, pressures of thousands of kg/cm² would be required to produce analogous effects for solid and liquid, if indeed they could be produced at all. With every extension of pressure range the probability of the existence of any such critical phenomena has become increasingly remote. Melting curves have now been followed up to 40,000 kg/cm²; a number of these are shown in Fig. 8. The melting curves of all substances have certain qualitative features in common, so that it is appropriate to speak, of *the* melting curve just as one may speak of *the* vaporization curve. In other respects, however, the situa-

tion with regard to melting is qualitatively different from that with regard to vaporization. In particular, all melting curves, that is, the curve of melting temperature against pressure, are concave toward the pressure axis with a curvature becoming less at higher pressures, and the curve of difference of volume between liquid and solid as a function of pressure is convex toward the pressure axis with a curvature decreasing with increasing pressure. No critical point has ever been observed in the experimental range. If there were such a point outside the range, the latent heats and the volume difference between liquid and solid would have to vanish at a common pressure and temperature. Extrapolation of the curves for latent heat and volume difference indicates that neither of them will vanish at any finite pressure or temperature, to say nothing of both vanishing at the same pressure and temperature. The probability at present seems overwhelming that there can be no critical point between liquid and solid, at least for the type of substance investigated up to now, which includes organic substances of various types and a few metals. The same line of argument rules out the existence of other such features on the melting curve as a maximum temperature or an asymptotic temperature. In general, the melting curve rises to indefinitely high temperatures with indefinitely increasing pressure but at a diminishing rate, the curve becoming more nearly linear.

It is possible to show thermodynamically that if a substance expands when it melts, its melting temperature must rise with increasing pressure, and, conversely, it falls. There are only three substances which belong to the latter category in the ordinary range, water, bismuth, and gallium. Consistent with thermodynamics the melting curves of these three substances are found to fall. Furthermore, the curvature increases and the curves drop more and more rapidly as pressure increases. Such a state of affairs apparently cannot continue indefinitely. Nature extricates itself from the dilemma by the «liquidating» of such abnormal substances. Above a certain pressure the lattices in which these substances initially crystallize become unstable, and the lattice collapses into another lattice. The new lattice has a volume so much less than the former lattice that the solid phase is now more dense than the liquid, and from here on the melting curve rises as for other substances. The collapse of the lattice occurs at a pressure of about 2,000 kg/cm^2 for water, at 12,000 for gallium, and at 25,000 for bismuth.

The phase changes of these three substances afford a particular example of polymorphism. The phase diagram of bismuth is shown in Fig. 9. Under pressure, polymorphism is a very common phenomenon; the number of in-

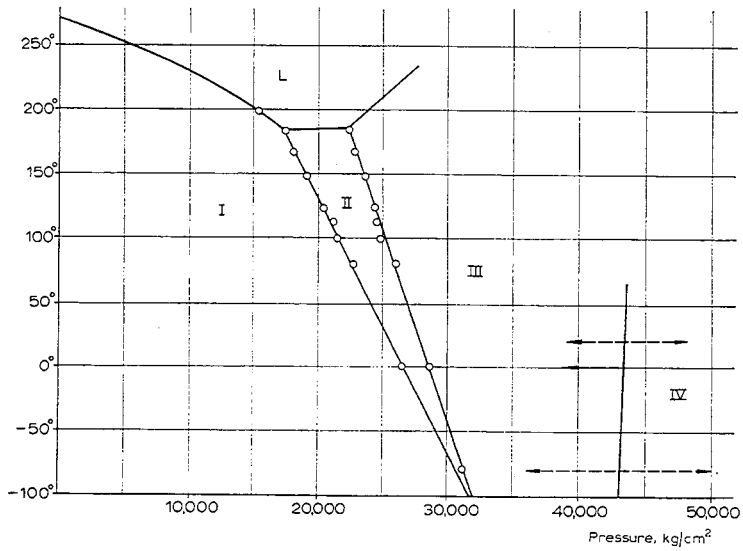


Fig. 9. The phase diagram of bismuth. The arrows on the transition line III–IV indicate the pressure limits within which the transition runs with increasing or decreasing pressure.

stances increases with increase in the experimental pressure range and with increasing sensitiveness in the methods for detecting small discontinuities of volume. In the range from room temperature to 200°C and up to pressures of 50,000 kg/cm², roughly one-third of the substances examined have proved to be polymorphic. In the much greater range of conditions encountered in the crust of the earth, the presumption seems to be that no substance exists in the lattice with which we are familiar under laboratory conditions, unless perhaps the lattice is of a particularly simple type. The importance of such a conclusion for geophysics is obvious.

The thermodynamics of a polymorphic phase change is the same as the thermodynamics of melting, but beyond that there is little resemblance between the two phenomena; there is no such thing as *a* polymorphic transition curve as there is *a* melting curve. There are only three falling melting curves, and these disappear at higher pressures; there are many falling transition curves, and they become increasingly numerous at higher pressures. Between 12,000 and 50,000, 41 per cent of the new transition curves are of the falling type. Transition curves may have horizontal or vertical tangents; melting curves have neither. Transition curves may have upward or downward curvature; melting curves are always concave downward. The dif-

ference of volume of two polymorphic phases may increase or decrease in the direction of increasing temperature along the transition line; the difference of volume between liquid and solid always decreases. The compressibility of the high-pressure phase may be greater or less than that of the low-pressure phase; the compressibility of the liquid is always greater than that of the solid. Substances are capable of existing in a number of polymorphic forms, and the complete mapping of the transition temperatures and pressures for all the forms may result in phase diagrams of great complication. Thus bismuth has six different phases; water, which has some striking analogies to bismuth, has seven phases. The most complicated phase diagram investigated to date is that of camphor, which has eleven phases.

There are only two generalizations with regard to polymorphic transitions that stand to date. The first is that critical points and continuous transitions between different polymorphic forms do not occur. If there were such points this would demand a continuous transition from one type of lattice to another, and this seems highly improbable, although perhaps not logically inconceivable. The second generalization is that transitions in the simple CsCl type of lattice in the direction of smaller volumes are not produced by pressure; this lattice seems to be of such a simplicity that it is not likely to be disturbed. This second generalization naturally rests on a much smaller number of examples than the first, and is correspondingly less secure.

We have so far been discussing transitions which are thermodynamically reversible; when pressure is released the original form is resumed. In addition to these reversible transitions, the existence of irreversible transitions is to be recognized, that is, of changes effected by pressure which remain permanent when they have once taken place. Two well-marked examples of this have been found. The first was phosphorus. If ordinary yellow phosphorus is exposed to pressure above 12,000 kg/cm^2 at temperatures above 200°C, it is permanently changed into a black solid much like graphite in appearance, and like it a conductor of electricity, although yellow phosphorus is a good insulator. This remained the only example for many years. Recently I have found that ordinary liquid CS_2 may similarly be changed permanently into a black solid at temperatures in the neighborhood of 200°C and by pressures of the order of 40,000 kg/cm^2. This black substance is definitely not a mixture of sulfur and carbon, which one might at first expect, but is apparently a unitary substance, truly a black solid form of carbon bisulfide. It has been suggested that the structure may be that of a single giant molecule like the known structure of SiO_2, which from the atomic point of view is very sim-

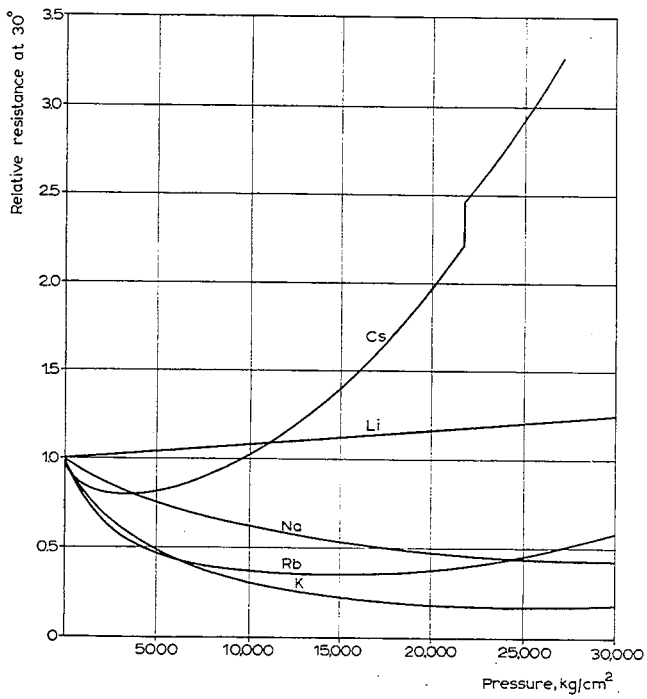


Fig. 10. The relative resistances of the alkali metals up to 30,000 kg/cm². The break in the curve for caesium is due to a polymorphic transition. Potassium has a very flat minimum near 23,000.

ilar. It is fascinating to speculate that there may be many other common substances which may be pushed by sufficiently high pressures over a potential hill of some kind permanently into some hitherto unknown form. Until we have theoretical understanding of these two known permanent transitions, we can not help attaching a certain reasonableness to the assumption of the possible existence of other such substances. In fact, there is experimental evidence that many other such transformations may be effected. In experiments in which I combined high shearing stresses with high hydrostatic pressure I have observed some cases of irreversible transitions to forms already known, and have also observed a large number of color changes, which are the indication of some sort of permanent change. It was not possible to establish whether new substances were formed under these conditions because the quantities of material involved were too small to permit satisfactory analysis.

We pass now to other sorts of pressure effects. Perhaps the simplest of these to measure are the effects of pressure on electrical resistance. Measurements

have been made at room temperature or higher up to 30,000 kg/cm² and at the temperature of liquid air to 7,000. At low temperatures there is a natural limit to the pressure range imposed by the freezing of the medium transmitting pressure, which in this case was gaseous nitrogen. Fig. 10 shows the effect of pressure on the alkali metals at room temperature up to 30,000 kg/cm².

In the first place, there is a specific effect of pressure on resistance; the pressure coefficient of resistance is in general of the order of magnitude of ten times greater than the volume compressibility. This would involve as a corollary that the effect of pressure on the resistance of a highly compressible metal is greater than on a metal of low compressibility. This is indeed true in general, but exceptions are common. The resistance of perhaps three-quarters of the metals decreases with increasing pressure; as is to be expected, the rate of decrease itself decreases with increasing pressure, that is, the curve of resistance against pressure is convex toward the pressure axis. On the other hand, there are several metals, of which lithium, strontium, and bismuth are examples, whose resistance increases under pressure. For these metals, surprisingly, there is a law of increasing returns, that is, the rate of increase of resistance itself increases with increasing pressure. This means that for these metals also the curve of resistance against pressure is concave upward. Finally, there are a few metals which combine both types of behavior, that is, the resistance initially decreases, then passes through a minimum, and then turns upward. Examples are caesium, rubidium, potassium, and barium. It would appear, therefore, that the upward curvature is common to all metals, and that all resistance curves may be regarded as pieces of one single curve, the only difference for different metals being that what might be called the intrinsic zero of pressure is differently situated with respect to atmospheric pressure for different metals.

Considerable success has been achieved by theoretical physicists in explaining theoretically the effect of pressure on resistance. As might be expected when effects departing so largely from linearity are concerned, we can recognize the presence of at least two different mechanisms working in opposite directions. There is in the first place an effect of pressure on what is the analogue of the electron free path in the old gas electron theory of metallic conduction. This is connected with the change of dimensions, and in general works in the direction of an increase of free path, that is, a decrease of resistance, with increasing pressure. In the second place, there is a rearrangement of the energy levels, and this may, when the energy bands are nearly

completely occupied, work in the direction of a change in the effective number of free electrons. Depending on the details of the atomic relations, this effect may be either an increase or a decrease. The calculations have been carried through approximately in a few simple cases. It turns out that the increase of resistance of lithium with pressure is due to the preponderating effect of a decrease in the effective number of free electrons.

The effect of pressure on the electrical resistance of single crystals is sometimes very complicated. If the crystal system is cubic, the material behaves electrically like an isotropic body, but if the system has lower symmetry, there may be differences in different directions. In antimony, for example, the sign of the pressure effect is different in different directions. There are directions in the crystal for which the resistance passes through a maximum with increasing pressure, whereas for other directions the resistance decreases with normal curvature.

The resistance of some semi-conductors may be decreased by such large amounts that they approach the absolute resistance characteristic of the metals. An early investigation in this field was made on selenium and silver sulfide by Montén in Uppsala. At higher pressures, tellurium approaches the properties of a metallic conductor under 30,000 kg/cm^2. Not only does the absolute value of the resistance drop to a characteristically low figure, but the temperature coefficient, which initially is negative, reverses sign under high pressure and becomes positive like that of the metals proper. Theory is as yet hardly in a position to explain these complicated effects, either in single crystals or in semi-conductors.

Closely related to the electrical conductivity of metals is their thermal conductivity; the relationship is expressed by the approximate equality of the Wiedemann-Franz ratio of electrical to thermal conductivity for all metals. Under pressure, thermal conductivity changes as well as electrical conductivity. It is much more difficult to measure than electrical conductivity, and satisfactory measurements have been made only for a few metals and those up to only 12,000 kg/cm^2. It appears that for these metals the Wiedemann-Franz ratio is approximately independent of pressure.

The effect of pressure on the thermal conductivity of liquids is much larger than on that of metals, and is much easier to measure. In general, the thermal conductivity increases for ordinary liquids under a pressure of 12,000 kg/cm^2 by a factor varying between 2 and 3. The effect on water is smaller; at 12,000 the increase for it is only 50 per cent. There is a close connection between the effect of pressure on thermal conductivity of normal liquids and

the effect of pressure on the velocity of sound in the liquid. That is, thermal conductivity in a liquid is primarily a mechanical affair; heat is transferred by microscopic mechanical waves travelling with the velocity determined in the conventional way by the compressibility. The small effect in water is associated with the small change in the compressibility of water brought about by pressure.

Another property of metals obviously related to electrical and thermal conductivity is the thermo-electric characteristics. These properties are also affected by pressure. In general, a metal under pressure behaves thermo-electrically differently from the same metal not under pressure, so that a thermocouple may be made with one branch of any ordinary metal and the other branch of the same metal exposed to hydrostatic pressure. Under a pressure of 12,000 kg/cm^2 the thermo-electric power of such couples may be as large

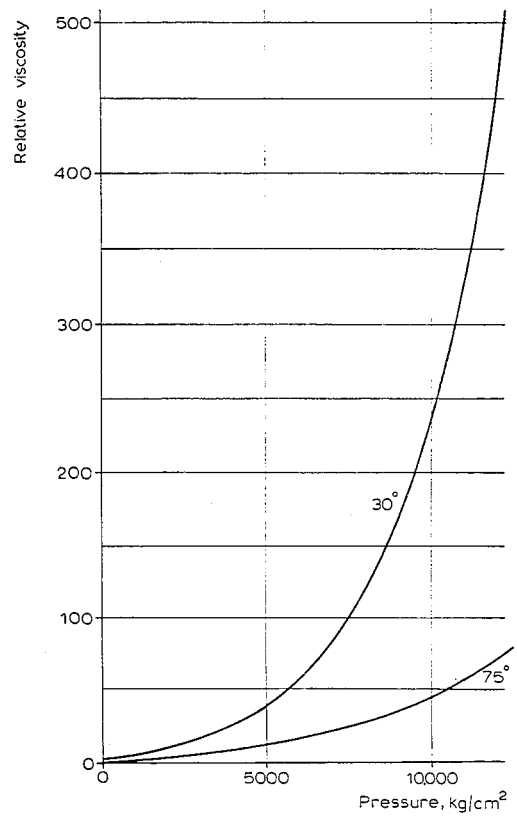


Fig. 11. The effect of pressure on the viscosity of isobutyl alcohol.

as that of ordinary couples composed of two entirely different metals. A number of such « pressure couples » have been investigated. The effects are complicated; there is not any universal rule with regard to the sign of the effect. There may be reversals of sign and large departures from linearity. No satisfactory theory of these effects has been formulated. At present one can only draw the conclusion that the interplay of several different mechanisms must be involved.

The largest of all the pressure effects studied to date is on the viscosity of liquids. In general, viscosity increases under pressure at a rate increasing rapidly with increasing pressure. The curve of viscosity against pressure usually rises exponentially with pressure, and sometimes more rapidly than exponentially. In Fig. 11 is shown the viscosity of isobutyl alcohol at 30° and 75° at pressures up to 12,000 kg/cm^2. The total rise may be by as much as a factor of 10^7 for a pressure increase of 10,000 kg/cm^2 (for eugenol). The rate of rise is definitely linked with the complication of the molecule, and is greater the more complicated the molecule. For the comparatively simple liquid water the rise of viscosity under 10,000 kg/cm^2 is by a factor of only 2 or 3 and for monatomic mercury by only 30 per cent. For methyl alcohol the increase is by a factor of 10, for propyl alcohol by a factor of 100, and for amyl alcohol by a factor of 1000. In the last few years the theoretical physicists have had considerable success in accounting for the effect of pressure on the viscosity of liquids.

Biography

Percy Williams Bridgman was born in Cambridge, Massachusetts, on April 21st, 1882. He received his early education in public schools in the nearby city of Newton until 1900 when he entered Harvard University. He graduated A.B. in 1904, A.M. in 1905 and was awarded his Ph.D. (Physics) in 1908 whe he joined the Faculty of the University. Bridgman was successively appointed Instructor (1910), Assistant Professor (1919), before becoming Hollis Professor of Mathematics and Natural Philosophy in 1926. He was appointed Higgins University Professor in 1950.

His researches concerning the effects of high pressures on materials and their thermodynamic behaviour commenced in 1905 and have continued throughout his career. He has carried out extensive investigations on the properties of matter at pressures up to 100,000 atmosphere including a study of the compressibility, electric and thermal conductivity, tensile strength and viscosity of more than 100 different compounds. He developed a method of packing which eliminated leak, and later introduced various methods of external support to pressure vessels as higher pressures were demanded. Bridgman has also contributed to crystallography, where he devised a method of growing single crystals; to the problems of electrical conduction in metals, where he discovered internal Peltier heat – a new electrical effect; and to the philosophy of modern physics. In the latter field, he is a strong supporter of the operational viewpoint, considering it meaningless to interpret physical concepts except as they are capable of observation.

Prof. Bridgman has contributed many papers to leading scientific journals and he is the author of *Dimensional Analysis* (1922), *The Logic of Modern Physics* (1927), *The Physics of High Pressure* (1931), *The Thermodynamics of Electrical Phenomena in Metals* (1934), *The Nature of Physical Theory* (1936), *The Intelligent Individual and Society* (1938), *The Nature of Thermodynamics* (1941), and, more recently, *Reflections of a Physicist.*

He has received honorary Doctor of Science degrees from Stevens Institute (1934), Harvard (1939), Brooklyn Polytechnic (1941), Princeton (1950), Paris (1950), and Yale (1951). He has received the Rumford Medal (Ameri-

can Academy of Arts and Sciences), the Cresson Medal (Franklin Institute), the Roozeboom Medal (Royal Academy of Sciences of the Netherlands), the Comstock Prize (National Academy of Sciences), and the New York Award of the Research Corporation. He was a member of the American Physical Society (President, 1942), the American Association for the Advancement of Science, the American Academy of Arts and Sciences, the American Philosophical Society, and the National Academy of Sciences. He was a Foreign Member of the Royal Society and Honorary Fellow of the Physical Society (London).

Bridgman married Olive Ware in 1912. Their daughter, Jane, was born in 1914, and their son, Robert Ware, in 1915.

Prof. Bridgman died in 1961.

Physics 1947

Sir EDWARD VICTOR APPLETON

«for his investigations of the physics of the upper atmosphere, especially for the discovery of the so-called Appleton layer»

Physics 1947

Presentation Speech by Professor E. Hulthén, member of the Nobel Committee for Physics

Your Majesty, Your Royal Highnesses, Ladies and Gentlemen.

On 12th December 1901, Marconi succeeded in establishing wireless communication between the Old World and the New. The way in which the wireless waves proved to follow the contour of the earth compelled the assumption that there must be an electrically conducting layer somewhere high up in the stratosphere. Thereby, it was thought, the linearly moving radio waves were thrown back towards the earth just as the rays of the sun are deviated so that its light can be observed long after the sun has passed below the horizon. According to Heaviside and Kennelly such a layer which throws back the radio waves and girdles the whole earth conceivably be due to the ionizing effects of the ultraviolet rays of the sun on the upper atmosphere. However, no conclusive proof of this was forthcoming even at the beginning of the 1920's.

At this time the rapid development of radio, especially in England and America, rendered the crowding of the ether so great that amateurs had to be relegated to wavelengths below 100 metres, while the kilometre-long waves were then considered to be most suitable for long-distance transmissions. The most immediate difficulties for the amateurs were to obtain transmitters with sufficient effect for the great distances. The competitions arranged for amateurs in radio signalling with short waves between America and Europe in the years 1921 and 1922 proved, however, that in sporadic cases communication could be obtained with surprisingly small transmitter effect.

It was here that Appleton's contributions now began to make themselves felt. By means of a brilliantly worked-out method, the so-called *frequency variation method*, in 1924 together with Barnett, he showed that there was interference between the direct radio waves along the ground and a wave reflected towards the layer in space, and that this so-called *Heaviside layer* is at about a height of 100 km. By means of a fine theoretical analysis of the penetration of the wave into the layer, Appleton was also able to investigate certain important detail phenomena which were observed in connection with these investigations. Thus the wave was refracted in the layer in the

same way as a ray of light entering an optically thin medium. Consequently radio waves will either be thrown back towards the earth or, if they are strong enough to penetrate to the middle of the layer where the ionization is greatest, they will run through it and out into space. According to the theory from the critical wavelength at which this takes place can be read off the ionization at different points of the stratosphere, often at great distances from the transmitter. During the last World War, stations with equipment for registering these conditions were established in different parts of the earth's surface. Information from these stations is of great importance for radio communication, as it affords guidance in the choice of the wavelength which should be selected in the case of radio communications between two places. But, as we shall see presently, these investigations are of considerable importance for many other things besides radio communication.

It was in the course of the study of these phenomena that Appleton found in 1927 that there must exist still another reflecting layer beyond the one mentioned previously, at a height of about 230 km. This so-called *Appleton layer* is still more exposed to the ultraviolet solar radiation than the underlying layer, or more correctly: the underlying layers, for a further couple of layers of more or less sporadic occurrence, have been proved in the course of time. Ionization is therefore more complete up there than in the underlying layers, and there is a greater power of reflection against the radio waves. Appleton has shown that during the day the new layer is divided into two components which again merge into one during the night. It is mainly the upper layer which, owing to its greater capacity of reflection, facilitates radio communications. In another respect also it differs from the lower layers. While in the latter the ionization constantly follows the changes in the ultraviolet solar radiation, the upper Appleton layer remains unchanged during the greater part of the night. This is explained by the great rarefaction prevailing at these heights, which retards the recombination of the ions of the air. The close correlation exhibited by the ionization of the lower layers with the changes in the ultraviolet solar radiation was convincingly shown by Appleton's observations between the time of the minimum of sunspots in 1934 and that of the maximum of sunspots in 1937. The increase in the ionization then amounted to 50 to 60%, corresponding to an increase in the intensity of the sun's ultraviolet radiation of 120 to 150%. This circumstance is so much the more remarkable as observations made on the surface of the earth during the same period show that the ultraviolet radiation there was practically unchanged. Thus the radio method has proved to be a means of

determining the actual radiation of the sun. Through the sunspots, which must be looked upon rather as holes or windows opening into the interior of the sun, we observe some of the mighty processes taking place there. Appleton has also found that the disturbances, which, in the form of short radio waves, emanate from these sunspots are equivalent to transmitter effects of millions of kilowatts.

It would carry us too far to give a detailed account here of all the discoveries and investigations which we have owed to Appleton during the last few years. I shall only dwell briefly upon one aspect of them. The echo methods which were developed by Appleton and his co-workers in the years before the World War must be looked upon as precursors of the *radar methods* which were so successfully employed by the Allies during the War for locating aeroplanes, submarines, etc. The ultrashort radar waves, 3–30 cm in length, will certainly be employed for many purposes in the immediate future, *inter alia* as an important auxiliary aid within *meteorology*. The direction of radar waves is not changed during their passage through ionized layers, but they are changed owing to the inhomogeneities arising owing to variations in the pressure and temperature of the air and its varying content of aqueous vapour. Thus the radar waves are reflected by showers, which can be detected at far distances and determined by the radar echo. During the World War radar methods were extensively used to locate the heat front and the cold front in distant low-pressure areas.

Finally, it may be mentioned that Appleton has carried out far-reaching investigations of the electric waves which are produced when the lightning strikes. With the help of specially equipped sounding stations, lightning discharges and thunderstorms which are 1,000 to 2,000 km away can be located, and the disturbances which affect radio reception he has found to be due to interplay between far-distant thunderstorms, especially on the equator.

Thanks to Appleton's contributions a new branch has been added to physical science, but not only that: the methods which he and his co-workers have perfected to investigate the atmosphere round the earth by means of radio waves have also become of immense importance for solving problems within other sciences, such as astronomy, geophysics and meteorology, and for radio technics.

Sir Edward Appleton. Electromagnetic waves are a subject of the greatest physical importance, and they are being increasingly applied in different fields of science. The first arguments on the existence of these waves were

advanced more than a hundred years ago by your countryman Michael Faraday, who was then searching for the relations between optical and electrical phenomena. His ideas were worked out in strict mathematical equations by James Clark Maxwell in 1873. The waves were finally discovered by the famous German physicist Heinrich Hertz in the early 1890's. Shortly afterwards their immense usefulness as radio waves was demonstrated by the Italian inventor Guglielmo Marconi.

Since then, electromagnetic waves have advanced victoriously in a multitude of sciences, giving rise, in the hands of men of genius, to scientific methods and instruments, among which I need only mention the electronic tube, based on the thermo-ionic laws, so thoroughly investigated by Sir Owen Richardson.

Now you have added a new link to this beautiful chain, applying the waves to the study of our own atmosphere. With the aid of these waves you have reached etherial regions never before attained by man. You have even taught us how to listen to the roar from eruptions in the sun and distant stars in the galaxy.

The usefulness of radar waves as applied to the solution of problems in meteorology has already been shown. The need for such a refined and instantaneous control of the unreliable and capricious conditions prevailing in the earth's atmosphere cannot be overestimated, especially in respect to the risks which still jeopardize aviation.

From Greek mythology we learn how Daedalus fastened a pair of wings to the shoulders of his son Icarus with wax. But Icarus flew too near the sun, and the wax melted, so that he fell into the sea and was drowned. Certainly the modern Icarus also needs to strengthen his wings in his flights.

On behalf of the Royal Swedish Academy of Sciences I congratulate you on your important discoveries, and I will now ask you to receive your Nobel Prize from the hands of His Majesty.

EDWARD V. APPLETON

The ionosphere

Nobel Lecture, December 12, 1947

In this lecture I wish to draw your attention to certain features of the electrical state of the higher reaches of the earth's atmosphere. This is a region which human beings have not yet visited, and so the information which we have accumulated about it is of an indirect character.

Now the most striking feature of the atmospheric air at high levels is that it is ionized, and for that reason the spherical shell surrounding the earth at the levels with which we are concerned is called the ionosphere. It has, of course, been suspected for many years that when there is an auroral display, yielding intense luminosity, the upper atmosphere must be strongly ionized like the gas in a Geissler tube. But I am not so much concerned today with irregular events of that kind, which occur mainly in high latitudes, as with the permanent shell of ionization which exists at all times and at all latitudes, even above the equator.

There were originally two lines of evidence which suggested that the upper atmosphere might be electrically conducting. In the first place Balfour Stewart, in 1882, put forward the hypothesis that the small daily rhythmic changes of the earth's magnetic field were due to the magnetic influence of electric currents flowing at high levels. Balfour Stewart pictured such currents as arising from electromotive forces generated by periodic movements of the electrically conducting layer across the earth's permanent magnetic field. The movements, he suggested, were largely tidal in character and therefore due to the gravitational influence of the sun and the moon.

The second indication of the possible existence of a conducting layer in the upper atmosphere came from the study of the long-distance propagation of radio waves. The successful communication established by Marconi between England and Newfoundland in 1901 prompted many theoretical studies of the bending of electric waves round a spherical earth. These mathematical investigations of radio-wave propagation round a conducting sphere showed conclusively that Marconi's results could not be explained in terms of wave diffraction alone. Some factor favouring radio transmission over long distances had evidently not been taken into account.

Suggestions as to the nature of this factor were fortunately to hand, for in 1902 Kennelly and Heaviside had independently pointed out that, if the upper atmosphere were an electrical conductor, its influence would be such as to guide the radio waves round the earth's curvature, energy being conserved between the two concentric conducting shells and so not lost in outer space.

The Kennelly-Heaviside theory did not, however, gain universal acceptance, for direct evidence of the existence of the conducting layer was lacking. Opponents of the theory, for example, sought to explain Marconi's results in terms of the refractive bending of the waves due to the stratification of the air and water vapour in the lower atmosphere near ground level.

During the 1914–1918 War, when I served as a radio officer in the British Corps of Royal Engineers, I became interested in the problems of radio propagation and the fading of radio signals. As a result, after the war, when I returned to Cambridge, I began to work on the subject, starting first to develop more accurate methods of radio-signal measurement. The initiation of broadcasting in Britain in 1922 greatly assisted these experiments, for powerful continuous wave senders became generally available for the first time. Measurements of received signal intensity, made at Cambridge on waves emitted by the London B.B.C. sender, showed that, whereas the signal strength was sensibly constant during the daytime, slight fading was experienced at night. A possible explanation was that such fading was due to interference effects between waves which had travelled straight along the ground, from sender to receiver, and waves which had travelled by an overhead route by way of reflection in the upper atmosphere.

We may picture such a state of affairs as shown in Fig. 1. Here we see that radio waves can travel from the sender to the receiver by two paths – one direct and one indirect. Now if there is a whole number of wavelengths in

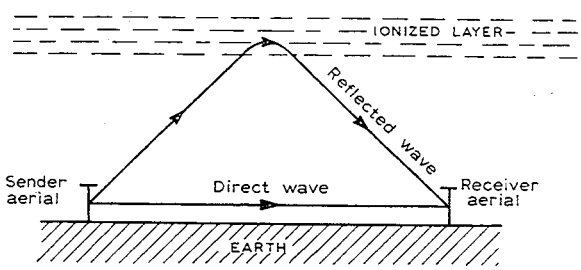


Fig. 1.

the path-difference between the ground and the atmospheric ray paths, there will be a maximum of radio-signal intensity at the receiver; while if the path-difference is equal to an odd number of half-wavelengths a minimum of signal will be experienced. Let us suppose now that the wavelength of the radiation emitted by the sender is slowly and continuously altered. This will produce a succession of maxima and minima of signal intensity at the receiver, and, if the number, n, of either is counted, and the initial and final wavelengths (λ_1 and λ_2 respectively) of the change are known, the difference in length, D, between the ground and atmospheric paths, may be found. We have, in fact, the relation:

$$n = \frac{D}{\lambda_1} - \frac{D}{\lambda_2} \qquad (1)$$

if D does not alter sensibly with wavelength. When the path-difference, D, is known, the equivalent height of the reflecting layer can be found by simple triangulation.

The first experiment of this type, using the variation of wavelength (or frequency) was carried out on December 11th, 1924, with the assistance of M. A. F. Barnett, my first research student. The B.B.C. sender at Bournemouth was used and the receiving station was established at Oxford. This experiment immediately yielded evidence of a sequence of artificially produced maxima and minima of received signal intensity. The estimated height of reflection was found to be about 90 km above the ground.

In another series of experiments, the angle of incidence of the reflected waves on the ground was measured. This was done by comparing the simultaneous signal variations on two receivers, one with a loop aerial and the other with a vertical antenna. These results also indicated the reception of downcoming waves from about the same level in the upper atmosphere. The two sets of observations therefore directly established the existence of the Kennelly-Heaviside Layer.

In the winter of 1926–1927, using experimental methods of the same type, I found that, before dawn, the ionization in the Kennelly-Heaviside Layer (« E Layer ») had been sufficiently reduced by recombination to permit of its penetration. Reflection, however, was found to take place at an upper layer which was richer in ionization and which I termed the « F Layer », the lower boundary of which was found to be situated at a level 230 km above the earth.

Since the positions of the two main reflecting layers were first established, experiments of the kind I have described have been continued and extended. It was soon found that the technique could be so improved as to permit the study of radio reflection of radio waves incident normally on the reflecting layers. This greatly simplified the interpretation of the results. In addition to using the frequency-modulation method of measuring the distance of the reflecting stratum, which has been described above, the elegant pulse-modulation method of making the same type of measurement, which Breit and Tuve had invented in 1925, was also adopted and developed. This method has proved the most powerful tool in ionospheric research. In its application in England, cathode-ray oscillograph delineation of the ground pulse and the subsequent echo pulses was also employed. It was by using a technique of this kind that the phenomenon of magneto-ionic splitting of echoes was discovered by G. Builder and myself. This confirmed, in a direct manner, what J. A. Ratcliffe and I had previously suspected from our experiments on the circular polarization of downcoming waves, that the ionosphere was a doubly-refracting medium due to the influence of the earth's magnetic field. The result indicated that free electrons, and not atomic or molecular ions, were the effective electrical particles in the ionosphere and paved the way for the development of the basic theory of a method of measuring electron densities in the ionosphere.

In the very early experiments on the ionosphere it was customary to use a constant frequency for the exploring waves and study the variation of equivalent height of reflection (h^1), as a function of time (t). In 1930, however, I suggested that perhaps more information could be obtained by studying the relation between height of reflection (h^1) and frequency (f), since, in this way, it would be possible to find the « critical » frequency of penetration for any layer. I had found, for example, that the « critical » penetration frequency for the E Layer, just before dawn, was about 0.75 Mc/s but that its value in summer daytime was about 3.0 Mc/s. It was from considerations such as these that the critical frequency method of measuring upper-atmospheric ionization was evolved. From a general theory of propagation of radio waves in a magneto-ionic medium I had found, in 1927, that the refractive index of such a medium became reduced to zero, permitting the reflection of radio waves at vertical incidence, when

$$N = \frac{\pi e^2}{m} f_o^2 \text{ (ordinary wave)} \qquad (2)$$

and

$$N = \frac{\pi e^2}{m}(f_x^2 - f_x f_H) \quad \text{(extraordinary wave)} \qquad (3)$$

for the conditions which usually obtain in practice in temperate latitudes. Here N is the electron density (electrons per cc) at the atmospheric level at which the refractive index becomes zero for an ordinary wave of frequency f_o and an extraordinary wave of frequency f_x. The quantities e and m are, respectively, the charge and the mass of the electron, while f_H is the gyro-angular-frequency (eH/mc), with which the electrons spiral round the lines of the earth's total magnetic force, H, at the level of reflection, c being the velocity of light.

Now, if, in Eqs. (2) and (3), the values of f_o or f_x refer to critical penetration, and therefore maximum, values of frequency, the corresponding values of N refer to maximum electron densities for the ionized layer which is just penetrated.

It will therefore be seen that, by finding the critical penetration frequency for either the ordinary ray or the extra-ordinary ray (which may be distinguished by their characteristic polarizations) it is possible to find the maximum electron density for any layer at any given time. For experimental convenience it is usual to employ the value of the ordinary-ray critical frequency f_o (see Eq. (2)), since that quantity is the easier one to determine experimentally.

(It should, however, be pointed out that, if both f_o and f_x are determined for the same conditions, it is possible to calculate f_H and so determine the value of H at the level of reflection. Using this method I have determined the value of the earth's total magnetic force H at the level of about 300 km, and shown that its value is approximately 10% less than its value at the ground.)

The first systematic experiments on the determination of the variation of the electron densities in the ionosphere were carried out in a 24-hour run on January 11–12, 1931, the E Layer being selected for study. It was then found that the E Layer maximum electron density starts to increase round about sunrise, reaches a maximum at noon and then wanes as the sun sets. Through the night, the ionization sinks to a low value, though there are often observed nocturnal sporadic increases of ionization which may possibly be due to meteoric dust. Later, the same critical frequency method was applied in the study of the F Layer. In this way there was inaugurated the long-term

study of the ionization in the various layers which has continued to the present time, when over 50 stations, using the critical-frequency method, are operating in different parts of the world.

Such continuous measurements of ionospheric densities, started in January 1931, in England, immediately showed a variation of noon ionization in sympathy with sunspot activity, which, in turn, indicated that the ultraviolet light from the sun, which is responsible for the electron production, varied substantially through the sunspot cycle. It was found, for example, that the E Layer ionization density was 50% greater in years of sunspot maximum than in years of sunspot minimum, indicating that the solar ultraviolet varied by as much as 125%, between the same two epochs. No such variation is to be noted in the heat and light we receive at ground level from the sun throughout the sunspot cycle.

Further work has confirmed the existence of a more weakly ionized region below the E Layer and which, in 1927, I had termed the «D Layer». The D Layer acts chiefly as an absorbing stratum for high-frequency radio waves, though the reflection of extremely low-frequency waves has been detected from it. It has also been found that the F Layer, especially under summer daytime conditions, tends to bifurcate into two overlapping strata, known as the F_1 and F_2 Layers. Experimentally we can therefore determine the critical penetration frequencies for the E, F_1, and F_2 Layers and so study the variation of their maximum ionization densities from hour to hour, from season to season, and from year to year. It has not yet been found experimentally possible to study the variation of ionization density of the D Layer in a similar way, but, with the assistance of W. R. Piggott, I have been able to show that measurements of ionospheric absorption indicate that the ionization in the D Layer varies in sympathy with sunspot number, in a manner somewhat similar to that found in the E and F_1 Layers.

The F_2 Layer, on the other hand, has been found to exhibit some remarkable anomalies, the full nature of which is only now becoming clear as a result of the world-wide study of the ionosphere conducted by the network of observing stations already mentioned. For example, in 1934–1935, working with R. Naismith I found that the ionization density in the F_2 Layer was actually less at summer noon than at winter noon, a result entirely at variance with the result for the D, E, and F_1 Layers, in which the ionization, as one would expect from theory, is greater in summer than in winter. It appears, in fact, that the F_2 Layer ionization is subject to factors additional to the normal solar control by variation of solar zenith distance. In this connec-

tion I have found, from a study of the results from many stations, that the F_2 Layer noon ionization density is, to a certain extent, controlled by magnetic latitude. A complete theoretical explanation of this phenomenon is still lacking.

Although, therefore, work on the systematic study of the ionosphere has now been in progress for over a quarter of a century it will be seen that the subject still presents us with unsolved problems, largely because of the different results obtained in different latitudes and longitudes. Ionospheric phenomena in auroral latitudes, for example, were the subject of study by an expedition from Great Britain to Tromsö in 1932–1933. There it was found that, under weak auroral and magnetic storm conditions, abnormal ionization was detected at night at E Layer levels, while, under strong auroral and magnetic storm conditions, extremely high absorption of the radio waves was experienced. No comparable studies of ionospheric conditions have yet been made in the Antarctic regions.

The work I have described was carried out with the object of exploring the electrical conditions in the higher atmosphere. It has yielded results which have shed light on many allied branches of science. In particular, the results disclosed have shown the essential correctness of the theory of Balfour Stewart concerning the origin of the rhythmic variations of terrestrial magnetism. For not only has it been shown that the upper atmosphere is electrical-conducting but it has been shown that the variation of that conductivity through the period of the sunspot cycle is of exactly the magnitude required to account for the variation of geo-magnetic changes with sunspot number. Moreover, from a careful study of layer heights the existence of large tidal movements, another essential feature of the Balfour Stewart theory, has also been demonstrated.

On the practical side of applications, ionospheric research has provided the basic ideas underlying the development of practical radiolocation of solid objects, for both the pulse-modulation and the frequency-modulation methods of measuring the distance of a reflecting surface by radio means have been used in the techniques of radar. Also, since we now have a fair understanding of the way in which the ionization in the reflecting layers varies through the day, through the season, and through the sunspot cycle, it is possible to forecast what I may call the «ionospheric weather» some time ahead. There has thus developed, on the practical side, the subject of «ionospheric forecasting» by which it is possible to forecast, say, three months ahead, the most suitable wavelengths for use at any time of the day, over

any distance of transmission, at any part of the world. In this way, scientific work conducted in the first instance with the object of exploring the wonders of the world around us, is now indicating how nation can speak unto nation with greater clarity and certainty.

Biography

Edward Victor Appleton was born in Bradford, England, on 6th September, 1892, the son of Peter and Mary Appleton. He received his early education at Hanson Grammar School. Bradford then took his B.A. degree in Natural Science at St. John's College, Cambridge, in 1913 and 1914, with physics for Part II. He won the Wiltshire Prize in 1913 and the Hutchinson Research Studentship in 1914, studying under Sir J. J. Thomson and Lord Rutherford. During the First World War he joined the West Riding Regiment, transferring later to the Royal Engineers. At the conclusion of hostilities he returned to Cambridge and took up research on radio waves.

Since 1919 Appleton has devoted himself to scientific problems in atmospheric physics, using mainly radio techniques. In 1920 he was appointed assistant demonstrator in experimental physics at the Cavendish Laboratory. Two years later he became sub-lector at Trinity College.

In 1924 Appleton was appointed Professor of Physics at London University and served there for twelve years, returning to Cambridge in 1936 to take the Chair of Natural Philosophy.

In the latter part of 1924 Appleton began a series of experiments which proved the existence of that layer in the upper atmosphere now called the ionosphere. With the co-operation of the British Broadcasting Corporation the Bournemouth transmitter shot waves up to the layer to see if they were reflected by it and came back. The experiment was entirely successful, for the reflection was proved. Moreover, by a slight change of wavelength it was possible to measure the time taken by the waves to travel to the upper atmosphere and back. The position of the reflecting layer was thus identified and its height (60 miles above ground) determined. The method used was what is now called « frequency-modulation radar ». The ionosphere was thus the first « object » detected by radiolocation, and this led to a great development of radio research and to a military invention of the greatest importance in World War II.

Further experiments which led to the possibility of round-the-world broadcasting were carried out and in 1926 he discovered a further atmos-

pheric layer 150 miles above ground, higher than the Heaviside Layer and electrically stronger. This layer, named the Appleton Layer after him, reflects short waves round the earth. Three years later Appleton made an expedition to Northern Norway for radio research, studying the Aurora Borealis and in 1931 he published the results of further research on determining the height of reflecting layers of the ionosphere, including the use of a transmitter that sent out « spurts » of radio energy, and the photography of the received echo-signals by cathode ray oscillography. In 1932 he was elected Vice-President of the American Institute of Radio Engineers.

When hostilities broke out in 1939 Appleton was appointed Secretary of the Department of Scientific and Industrial Research – the senior British Government post concerned with physical science.

Researches into the atmospheric layers and cathode ray oscillography were developed for aircraft detection when Sir Robert Watson-Watt and his group of scientists, working on Appleton's findings, brought Britain's secret weapon to perfection. Commonwealth researchers working with Appleton in Britain all became leaders in the development of radiolocation in their home countries and Sir Robert Watson-Watt has stated that, but for Appleton's scientific work, radar would have come too late to have been of decisive use in the Battle of Britain. Appleton was knighted in 1941, being created K.C.B., and he was a member of the Scientific Advisory Committee of the War Cabinet which, in 1941, advised the Government that the manufacture of an atomic bomb was feasible. Later, under Sir John Anderson, and as technical head of the Department of Scientific and Industrial Research, he assumed administrative control of all British work on the subject. He paid a visit to the United States and Canada in 1943 to arrange details of collaboration between American and British scientists. He continued research work even during this arduous period and has demonstrated that ionospheric reflecting power varies with sunspot activities. Also, working with Dr. J. S. Hey of the Ministry of Supply, he discovered that sunspots are powerful emitters of short radio waves. An important result of Appleton's work has been the establishment of a system of ionospheric forecasts, in which more than 40 stations all over the world co-operate, enabling the production of the most suitable wavelengths for communication over any particular radio circuit.

In 1947, the year in which he received the Nobel Prize for Physics, he was also awarded the highest civilian decoration of the United States – the Medal of Merit – and was made an Officer of the French Legion of Honour. He

was also awarded the Norwegian Cross of Freedom for his war work. Appleton's work has been recognized by India, Norway and Denmark, and in 1948 he was appointed by the Pope to the Pontificial Academy of Science. He received the Albert Medal of the Royal Society of Arts, in 1950, for outstanding services to science and industrial research and was elected President of the British Association for the Advancement of Science for the Liverpool meeting in 1953. He has been Chairman of the British National Committee for Radio-Telegraphy and Honorary President of the International Scientific Radio Union. During the International Geophysical Year 1957–1958 he played an active part in the world planning of radio experiments as Chairman of the International Geophysical Year Committee of the International Scientific Radio Union, and continues to remain a scientific research worker. He is now engaged on the interpretation of I.G.Y. ionospheric measurements on a global basis.

In 1956 Sir Edward gave the Reith Lectures of the B.B.C. on « Science and the Nation ». Recent awards made to him have been the Gunning Victoria Jubilee Prize of the Royal Society, Edinburgh, in 1960, and the Medal of Honour of the Institute of Radio Engineers of America in 1962.

In 1915 Appleton married Jessie, daughter of the Rev. J. Longson, and they have two daughters.

Physics 1948

PATRICK MAYNARD STUART BLACKETT

«for his development of the Wilson cloud chamber method, and his discoveries therewith in the fields of nuclear physics and cosmic radiation»

Physics 1948

Presentation Speech by Professor G. Ising, member of the Nobel Committee for Physics

Royal Highnesses, Ladies and Gentlemen.

According to the statutes of the Nobel Foundation, the Nobel Prize for Physics may be awarded for « discovery or invention in the field of physics ». The Royal Swedish Academy of Sciences in awarding this years' prize to Professor P. M. S. Blackett of Manchester, for his development of the Wilson method and his discoveries, made by this method, in nuclear physics and on cosmic radiation, indicates by the very wording of the award, that its decision is motivated on *both* the grounds mentioned in the statutes. Particular weight may perhaps, in this case, be laid on the discoveries made, but these only became possible by Blackett's development of the method and the apparatus.

Experimental research on the different kinds of rays appearing in nuclear physics has always been based to a great extent on the power of an electrically charged atomic particle, when moving at high speed, to *ionize* the gas through which it passes, i.e. to split a number of gas molecules along its path into positive and negative ions. Thus, one is able to *count* the number of particles by means of the Geiger-counter tube; such a counter being a special, very sensitive kind of ionization chamber, in which even a few ions produced by the ray are sufficient to release a short-lived discharge by an avalanche-like process.

But the whole course of the particle appears infinitely more clearly by the method invented by C. T. R. Wilson in 1911 and named after him. The radiation is allowed to enter an expansion-chamber, containing a gas saturated with water vapour. A sudden expansion of the chamber cools the gas, and cloud-drops are then formed instantly around the ions produced along the tracks of the particles. By suitable illumination these tracks can be made to stand out clearly as if they had been described by luminous projectiles. The « Altmeister » of modern nuclear physics, Lord Rutherford, once called the Wilson chamber « the most original and wonderful instrument in scientific history ».

But still, the immense value of the Wilson method for research purposes

did not become really apparent until the early twenties, and the credit for this changed attitude was largely due to the work of Blackett, who has ever since been the leading man in the development of the method. Before 1932 his work dealt chiefly with the *heavy* particles, appearing in radioactive radiations. In 1925, he obtained the first photographs ever taken of a nuclear disruption, namely the disruption of a nitrogen nucleus by an alpha particle of high velocity; the photographs clarified quite definitely the main features of the process. In this investigation and others from the same period he also verified, by accurate measurements, that the course of a collision between atomic nuclei always follows the classical laws of conservation of momentum and energy, provided the energy value of mass, as given by the theory of relativity, is also taken into account. These two laws, together with the conservation law of electricity, i.e. that positive and negative electricity are always produced together in equal amounts, form a set of three fundamental principles of general validity.

Blackett was soon to give to these principles an unexpectedly rich content by new experimental discoveries. In 1932 namely, he turned his interest to the cosmic rays, which at sea level are mainly vertical. The Wilson cloud chamber had already begun to be used at different places for the study of these rays, but with very low efficiency, as only about every twentieth random photograph showed the track of a cosmic ray. This was due to the fact that the rays are disperse both in space and time, and they must pass through the chamber only about a hundredth of a second before or after the moment of expansion, if they are to give a sharp track. Nevertheless, Anderson had at the time succeeded in obtaining a few photographs, showing the temporary existence of free *positive* electrons. These electrons, on account of their strong tendency to fuse with negative ones, seemed to exist free in a space filled with matter, only as long as they move at a great speed.

Together with his collaborator Occhialini, Blackett now developed an automatic Wilson apparatus, in which the cosmic rays could photograph themselves: the moment of expansion was determined by two Geiger counters, placed one above and the other below the chamber and connected to a quick electrical relay in such a way, that the mechanism of the cloud chamber was released only when *simultaneous* discharges occurred in both counters, i.e. when a cosmic ray had passed through them both and thus in all probability also through the cloud chamber between them. In this way the efficiency of the Wilson chamber was multiplied many times over, and the method became of extreme importance in cosmic ray research.

Immediately after completing this apparatus, Blackett and Occhialini discovered, in cosmic radiation, positive and negative electrons appearing in *pairs*; their tracks were deflected in opposite directions by a superposed magnetic field and they seemed to start from some common origin, often situated in the wall of the chamber. Sometimes such tracks appeared in great numbers, whole « gerbes », on the same photographic plate, demonstrating the existence in the cosmic radiation of veritable « showers » of positive and negative electrons. Shortly afterwards they established, in collaboration with Chadwick, that electron pairs are also produced by hard gamma rays, i.e. by the radiation of ultrashort wavelength emitted by certain radioactive substances; here the energy relations could be studied more closely than in the case of cosmic rays.

I shall try to give an idea of the great importance of these experimental results, even beyond the fact that they established irrefutably the existence of positive electrons. The discovery of the *pair* creation of electrons led, on the theoretical side, to the acceptance of two fundamental radiation processes of a reverse nature, which may be called *transmutation of light into matter* (represented by electron pairs) *and vice versa*. These processes take place within the framework of the three fundamental principles, just mentioned, regarding the conservation of momentum, energy and electricity: a quantum of light passing close to an atomic nucleus, may thus be transformed into a pair of electrons; but this is possible only if its energy at least equals the sum of the energy values of the two electronic masses. Since the rest mass of each electron corresponds to $\frac{1}{2}$ million electron volts, the light must possess a frequency at least corresponding to 1 million electron volts. If there is an excess of energy (i.e. if the frequency of the light is still higher), this excess will appear as the kinetic energy of the two electrons created. Reversely, the meeting of two slow electrons, opposite in sign, results in their fusion and *annihilation* as material particles; in this process two light quanta, each of $\frac{1}{2}$ million electron volts, are formed; these fly out from the point of encounter in opposite directions, so that the total momentum remains about zero (for even light possesses a momentum directed along the ray).

Blackett and Occhialini immediately drew these conclusions from their experiments and were guided in so doing by the earlier mathematical electron theory elaborated by Dirac on the quantum basis. The existence of the « annihilation radiation » was shortly afterwards established experimentally by Thibaud and Joliot.

These fascinating variations in the appearance of energy, which sometimes manifests itself as light, sometimes as matter, have stimulated the distinguished French physicist Auger to exclaim enthusiastically, in a monograph on cosmic radiation: « Who has said that there is no poetry in modern, exact and complicated science? Consider only the twin-birth of two quick and lively electrons of both kinds when an overenergetic light quantum brushes too closely against an atom of matter! And think of their death together when, tired out and slow, they meet once again and fuse, sending out into space as their last breath two identical grains of light, which fly off carrying their souls of energy! » (As a memory aid Auger's metaphor is excellent; its poetical value is perhaps open to dispute.)

In the late thirties, Blackett continued his researches on the cosmic radiation and, using a still further improved Wilson apparatus, made extensive accurate measurements concerning the momentum distribution, absorbability, etc. of this radiation. By means of a new optical method he was able to measure extremely feeble curvature of the tracks, corresponding to electronic energies up to 20 milliard electron volts.

Professor Blackett. In recognition of your outstanding contributions to science, The Royal Swedish Academy of Sciences has awarded to you this year's Nobel Prize for Physics for your development of the Wilson method and your discoveries, made by this method, in nuclear physics and on cosmic radiation.

In my speech, I have tried to sketch a few, and only a few, of your achievements and, more particularly, to give an idea of the fundamental importance of the discovery of pair creation. To me has been granted the privilege of conferring upon you the congratulations of the Academy and of inviting you now to receive your Nobel Prize from the hands of His Royal Highness the Crown Prince.

PATRICK M. S. BLACKETT

Cloud chamber researches in nuclear physics and cosmic radiation

Nobel Lecture, December 13, 1948

The experimental researches with which I have been occupied during the 24 years of my career as a physicist have been mainly concerned with the use of Wilson's cloud chamber for the purpose of learning more about the intimate processes of interaction of the sub-atomic particles. On 12th December, 1926, C. T. R. Wilson gave his Nobel Lecture entitled « On the cloud method of making visible ions and the tracks of ionizing particles », and described in it how, after a long series of researches starting in 1895, he developed in 1912 this exquisite physical method. Some here will probably remember that C. T. R. Wilson was originally drawn to investigate the condensation of water drops in moist air through the experience of watching the « wonderful optical phenomena shown when the sun shone on the clouds » surrounding his Scottish hilltops. I, like all the other workers with the cloud chamber, the world over, are indebted more than we can express to his shy but enduring genius.

In 1919, Sir Ernest Rutherford made one of his (very numerous) epoch-making discoveries. He found that the nuclei of certain light elements, of which nitrogen was a conspicuous example, could be disintegrated by the impact of fast alpha particles from radioactive sources, and in the process very fast protons were emitted. What actually happened during the collision between the alpha particle and the nitrogen nucleus could not, however, be determined by the scintillation method then in use. What was more natural than for Rutherford to look to the Wilson cloud method to reveal the finer details of this newly discovered process. The research worker chosen to carry out this work was a Japanese physicist Shimizu, then working at the Cavendish Laboratory, Cambridge, to which Rutherford had recently migrated from Manchester. Shimizu built a small cloud chamber and camera to take a large number of photographs of the tracks of alpha particles in nitrogen with the hope of finding some showing the rare disintegration processes. Unfortunately Shimizu had to return unexpectedly to Japan with the work hardly started. Rutherford's choice of someone to continue Shimizu's work

fell on me – then in 1921 a newly graduated student of physics. Provided by Rutherford with so fine a problem, by C. T. R. Wilson with so powerful a method, and by Nature with a liking for mechanical gadgets, I fell with a will to the problem of photographing some half million alpha-ray tracks.

Shimizu's cloud chamber was improved and made fully automatic, taking a photograph every 15 seconds – this rapid rate was only possible because of its small size, 6.0 cm diameter by 1.0 cm deep. The first task was clearly to study the forked tracks due to the normal collisions of alpha particles with oxygen, hydrogen, and helium atoms (Figs.1, 2, and 3), so as to verify that the normal collisions were truly elastic – that is, that no energy was lost in the process. If M and m are the masses of the alpha particle and nucleus, φ and θ the angle of deflection of the alpha particle and the angle of projection of the nucleus, then the assumption that energy and momentum are conserved during the collision leads to the relation

$$\frac{M}{m} = \frac{\sin(2\theta + \varphi)}{\sin \varphi}$$

Since θ and φ can be determined from the photograph, the mass ratio can be calculated. If for some track this is found to agree with the known ratio of the masses, then we conclude that the collision is elastic.

The following table shows the results of measurement of collisions with the nuclei of oxygen, hydrogen and helium, and show that, within the experimental error, the collisions were elastic.

Recoil atom	φ	θ	$m_{calc.}$	$m_?$
Oxygen	76°6′	45°12′	16.72	16.00
Hydrogen	9°21′	65°39′	1.024	1.008
Helium	45°49′	43°56′	4.032	4.00

The study of these forked tracks was one of the first quantitative investigations of the dynamics of *single* collisions of sub-atomic particles.

In order to calculate the angles φ and θ from the two photographs taken from directions at right angles, it was necessary to work out a geometrical method, which took into account the fact that a photograph of an object does not represent an orthogonal projection of the object but a conical projection, that is a projection through a point on to a plane. Only by such a method was it possible to obtain the necessary accuracy. An indication of the

accuracy of the angular measurements later achieved is shown by some results obtained in collaboration with E. P. Hudson in 1927 and with D. S. Lees in 1932. Measurements of 16 forked tracks due to the collision of alpha par-

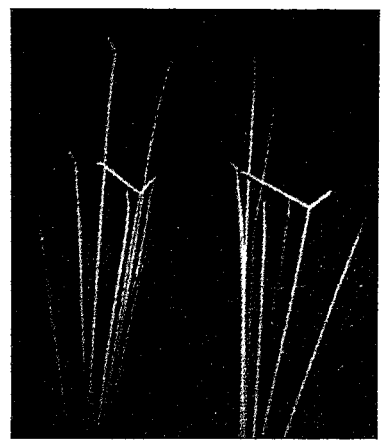


Fig. 1. Elastic collision of alpha particle with an oxygen nucleus.

Fig. 2. Elastic collision of alpha particle with a hydrogen nucleus.

ticles with hydrogen nuclei were made. The mean value of the mass ratio as calculated from the measured angles was 0.2531 ± 0.011, which differs from the known ratio 0.2517 by only a little more than the probable error. The probable error of a single angle measurement was estimated from these results to be 13′ of arc. The average energy, if any, lost in the collision must have been less than 1%.

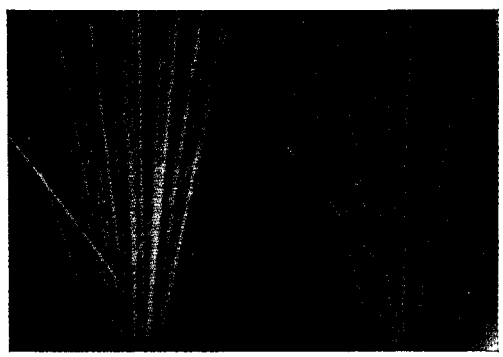


Fig. 3. Elastic collision of alpha particle with a helium nucleus.

Returning to the earlier period (1921–1924), a detailed study was also made of the relation between the range of a recoil nucleus and its velocity, the latter being calculated from the angles of the collision and the initial or final velocity of the alpha particle, assuming the collision to be elastic. The nuclei studied were those of hydrogen, helium, nitrogen, and argon. It was found that the range in air of a nucleus of mass m and atomic number z was approximately of the form

$$R \propto m z^{-1/2} \mathrm{f}(v)$$

where $\mathrm{f}(v)$ was roughly proportional to $v^{3/2}$. This relation was of importance in order to aid the identification of the recoil particles emerging from abnormal collisions.

This preliminary work done, production was started in earnest in 1924 and 23,000 photographs were taken within a few months. With an average of 18 tracks a photograph these gave over 400,000 tracks, each of which had to be scrutinized for anomalous behaviour. On some days when the apparatus worked well, as many as 1,200 photographs were taken. Eight forked tracks were found which had a quite different appearance from those showing

normal elastic collision, and these were readily identified as the sought for transmutation of nitrogen. Typical photographs are shown in Figs. 4 and 5.

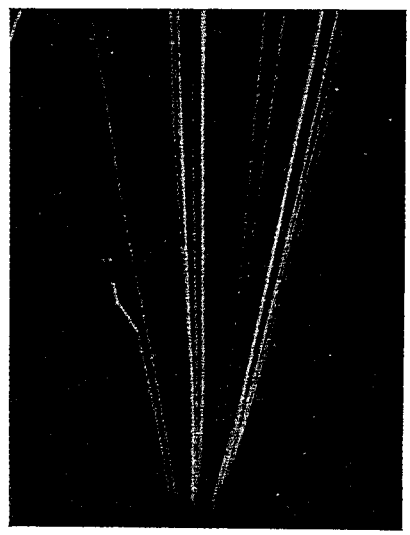


Fig. 4. Transmutation of nitrogen. One of the first photographs showing the capture of an alpha particle by a nitrogen nucleus with emission of a proton. The thin track moving to the right is the proton, and the short thick track to the left is due to the newly created ^{17}O nucleus. The alpha rays are from Thorium B + C (1925).

Rutherford's original experiments, using the scintillation technique, were only capable of proving that when an alpha particle struck a nitrogen nucleus a fast proton occasionally was ejected, but they were not able to reveal what happened to the alpha particle after the collision. There were two possibilities. The alpha particle might leave the nucleus again as a free particle, or it might be captured, so forming a heavier nucleus. In the former case, one would expect to find a forked track showing the track of the incident alpha particle, with three emergent tracks due to the alpha particle, the ejected proton, and the recoil nucleus. In the latter case one would find only two tracks, that of the proton and the recoil nucleus. The eight anomalous tracks all showed only two emergent particles, so proving that the assumed «disintegration» of nitrogen by alpha particles was in reality an «integration» process. Applying the principle of conservations of charge and mass, it was immediately deduced that the new nucleus formed must be a heavy isotope of oxygen $^{17}_{8}O$; the nuclear reaction being

$$^{4}_{2}\mathrm{He} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{1}_{1}\mathrm{H} + {}^{17}_{8}\mathrm{O}$$

At the time of these experiments this isotope of oxygen was not known, but shortly afterwards it was discovered by the analysis of band spectra.

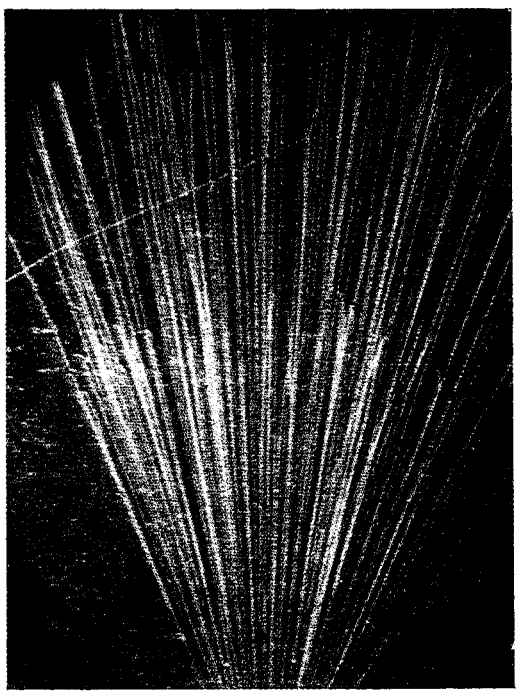


Fig. 5. Transmutation of nitrogen. The range of the alpha particle before the collision is only 2.4 cm and the range of the ejected proton, moving to the left and slightly backward, is only about 3.5 cm. The end of the 4.8 cm alpha particles from Thorium B + C can be seen near the middle of the photograph. (Blackett and Lees, 1932)

Since the ranges of the ejected protons were in all cases much larger than the size of the cloud chamber, it was not possible to determine directly their range or energy. However, the lengths of the tracks of the recoiling ^{17}O nucleus were readily measurable, and could be compared with that expected from the momenta of the particles, calculated on the assumption that momentum but not energy were conserved during the collision. The relation between the range and the momentum of an ^{17}O nucleus was not, of course, known directly, but could be estimated by interpolation, using the data for other common nuclei H, He, C, N, ^{16}O, and A, which, as has already been

explained, had previously obtained from the analysis of elastic collisions. In this way it was shown that the range of the recoil tracks was in good agreement with the calculated value for a mass of 17 and an atomic number of 8. Again, assuming the conservation of momentum, the sum of the energies of the two particles after the collision could be calculated, and was found to be on the average about 20% less than the energy of the incident alpha particle. The collision process was therefore an endothermic one – that is energy was absorbed in the process – so that the sum of the masses of the final products, ${}_{1}^{1}H$ and ${}_{8}^{17}O$, was somewhat larger than the sum of the masses of the original particles ${}_{2}^{4}He$ and ${}_{8}^{16}O$.

These experiments gave for the first time detailed knowledge of what is now known to be a typical nuclear transformation process. Owing to the laborious nature of the task of photographing the collisions of natural alpha particles with nuclei, not very much subsequent work has been carried out with this method. But with the discovery in 1932 of the neutron by Chadwick and of the disintegration of nuclei by artificially accelerated particles by Cockcroft and Walton, very many nuclear transformations have been studied in many laboratories by the use of the cloud chamber. In recent years the use of special photographic emulsions to record the tracks of nuclear particles, first used successfully by Blau and Wambacher, and later most fertilely exploited particularly at Bristol by Powell, Occhialini and their coworkers, has made possible the study of many types of nuclear collision processes with much greater facility than can be achieved with the cloud chamber.

After the work was completed, a larger automatic chamber of 16 cm in diameter was constructed and an attempt was made with D. S. Lees to photograph the disintegration of argon; 750,000 tracks were photographed on some 1,200 photographs, but no case of an argon disintegration was found. A further 350,000 tracks in nitrogen gave four more nitrogen disintegration processes, one of which was striking in that it was caused by an alpha particle of relatively low range (2.4 cm in air) and produced a proton track of only about 3.5 cm range (Fig.5).

In 1930 Mott had predicted by means of wave mechanics that the scattering of identical nuclear particles should differ markedly from that given by the inverse-square law. Theory showed that interference effects should occur somewhat analogous to the scattering of light by small particles. Chadwick had verified Mott's conclusion using fast alpha particles detected by scintillations. In collaboration with F. C. Champion, the scattering of

slow alpha particles with helium nuclei was studied using the automatic cloud chamber and a striking verification of Mott's theory was achieved.

In the autumn of 1931 in collaboration with G. P. S. Occhialini, I started to study the energetic particles found in cosmic rays by means of the cloud method. About 4 years previously Skobeltzyn in Leningrad had investigated the beta rays from radioactive sources using a cloud chamber in a magnetic field of 1,500 gauss. On some of the photographs he noticed a few tracks with very little curvature, indicating an energy over 20 MeV, that is much higher than any known beta ray. He identified these tracks with the particles responsible for the « Ultrastrahlung » or « cosmic rays », whose origin outside the earth's atmosphere had first been demonstrated in 1912 by the balloon flights of Hess and which had subsequently been much studied with ionization chambers by Millikan, Kolhörster, Regener, Hoffman, and others.

Skobeltzyn noticed also that these energetic particles occasionally occurred in small groups of 2, 3, or 4 rays, apparently diverging from a point somewhere near the chamber.

Skobeltzyn's work was followed up by Kunze in Kiel, and by Anderson in Pasadena. By using much larger magnetic fields up to 18,000 gauss, the energy spectrum of the particles was shown by these workers to extend to at least 5,000 MeV, and it was found that roughly half the particles were positively, and half negatively charged. The occasional association of particles was again noticed, particularly by Anderson.

The method used, that of making an expansion of a cloud chamber at a random time and taking the chance that one of the rare cosmic rays would cross the chamber during the short time of sensitivity – generally less than $\frac{1}{4}$ second – was much consuming of time and photographic film, since in a small chamber only some 2% to 5% of photographs showed cosmic ray tracks.

Occhialini and I set about, therefore, the devising of a method of making cosmic rays take their own photographs, using the recently developed « Geiger-Müller counters » as detectors of the rays.

Bothe and Rossi had shown that two Geiger counters placed near each other gave a considerable number of simultaneous discharges, called coincidences, which indicated in general the passage of a single cosmic ray through both counters. Rossi devised a neat valve circuit by which such coincidences could easily be recorded.

Occhialini and I decided to place Geiger counters above and below a ver-

tical cloud chamber, so that any ray passing through the two counters would also pass through the chamber. By a relay mechanism, the electric impulse from the coincident discharge of the counter was made to actuate the expansion of the cloud chamber, which was made so rapid that the ions produced by the ray had no time to diffuse much before the expansion was complete. The chamber was placed in a water-cooled solenoid giving 3,000 gauss. Having made the apparatus ready, one waited for a cosmic ray to arrive and take its own photograph. Instead of a small fraction of photograph showing a cosmic ray track, as when using the method of random expansion, the counter-controlled chamber yielded a cosmic ray track on 80% of the photographs. The first photographs by this new method were made in the early summer of 1932.

In the autumn of the same year, Anderson working with a normal chamber taking photographs at random, reported the finding of a track which he interpreted as showing the existence of a new particle – the positive electron.

The track described by Anderson traversed a lead plate in the centre of the chamber and revealed the direction of motion of the particle by the difference of curvature on the two sides. From the direction of motion and the direction of the magnetic field, the charge was proved positive. From the range and ionization, the mass could be proved to be much less than that of a proton. Anderson thus identified it as a new particle, the positive electron or positron.

During the late autumn of 1932, Occhialini and I, using our new counter-controlled cloud method, accumulated some 700 photographs of cosmic rays, among which groups of associated rays were so striking a feature as to constitute a new phenomenon and to deserve a name. From their appearance they came to be known as «showers» of cosmic ray particles. As many as 23 particles were found on a single photograph, diverging from a region over the chamber. Roughly half the rays were due to positively charged and half to negatively charged particles. From their ionization and range, the masses of the positive particles was evidently not much different from that of negative electrons. So not only was Anderson's discovery of the positive electron further confirmed by a wealth of evidence, but it was proved that the newly discovered particles occurred mainly in showers along with approximately an equal number of negative electrons. This fact of the rough equality of numbers of positive and negative electrons, and the certainty that the former do not exist as a normal constituent of matter on the earth, led us inevitably to conclude that the positive electrons were born together

Fig. 6. Cosmic ray shower. One of the first photographs of a large shower of cosmic ray particles. Some 16 particles, about half positive and half negative, diverge from a region over the chamber. This shower was interpreted as showing the birth of a number of pairs of positive and negative electrons. The counter-controlled cloud chamber was in a field of 3,000 gauss. (Blackett and Occhialini, 1933)

in collision processes initiated by high-energy cosmic rays. The energy required to produce such a pair is found from Einstein's famous equation to be $2\,mc^2 \cong 1$ MeV. So was demonstrated experimentally for the first time the transformation of radiation into matter.

The fate of the positrons was discussed in relation to Dirac's theory of holes. On this theory a positive electron was envisaged as a «hole» in a sea consisting of an infinite number of negative electrons in states of negative kinetic energy. Dirac's theory predicted that a positive electron would disappear by uniting with a negative electron to form one or more quanta. Occhialini and I suggested that the anomalous absorption of hard gamma rays by nuclei might be a result of the process of pair production, and that the observed re-emission of softer radiation might represent the emission of two 0.5 MeV quanta resulting from the annihilation of a positive and negative electron. Subsequent work has confirmed this suggestion.

This work was described in a paper which appeared in March 1933. Some of the photographs from the paper are reproduced here (Figs. 6–8). These represent the first published photographs showing positive electrons, as An-

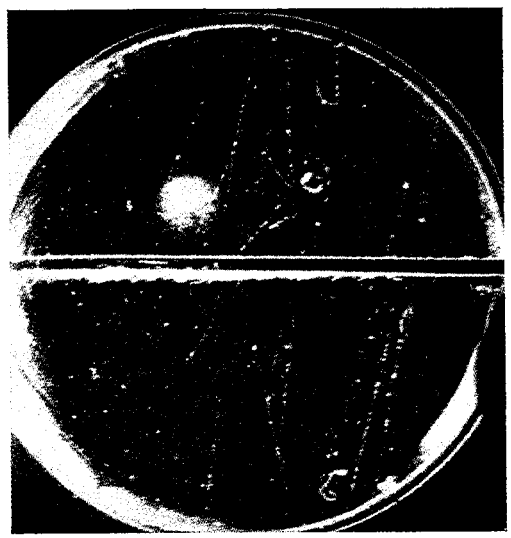


Fig. 7. Cosmic ray shower. Some 23 particles cross the chamber. Several radiant points can be detected above the chamber and also in the lead plate. H = 2,000 gauss. (Blackett and Occhialini, 1933)

derson's very beautiful photograph, though taken six months earlier, was not published till shortly afterwards.

The photographs showed clearly that some form of non-ionizing radiation must play an essential part in the formation of the showers, and that the mean range in lead of these radiations, which were assumed to be either photons or neutrons, must be quite small. Subsequent theoretical work by Heitler, Bethe, Bhabha and others gave a full account of these showers as due to a cascade process, consisting of the alternate emission of collision radiation by fast electrons and positrons, and the subsequent absorption of the latter by pair production.

As soon as the presence of positive electrons in cosmic rays was fully established, experiments were undertaken in collaboration with Occhialini and Chadwick to see if they were formed when hard gamma rays from radioactive sources were absorbed by matter. This was found to be the case when the energy of the rays was considerably above 1 MeV. One of the photographs of pair production by gamma rays is shown in Fig. 9.

It is interesting to note that the development of the counter-controlled cloud chamber method, not only attained the original objective of achieving much economy in both time and film, but proved to have the quite unex-

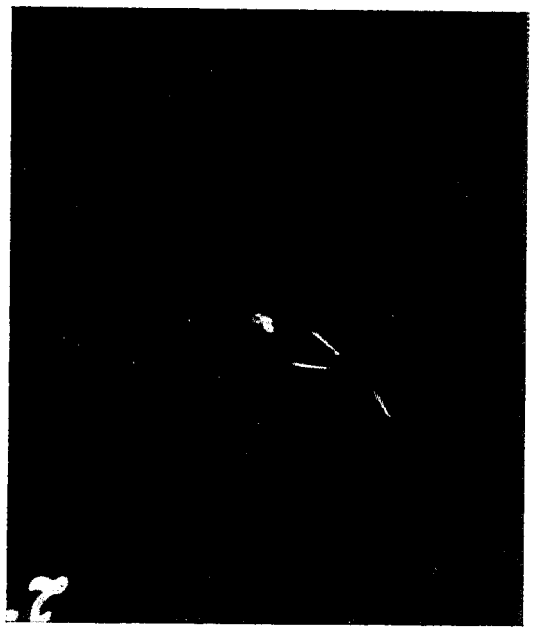


Fig. 8. Nuclear explosion produced by cosmic rays. Three heavily ionizing particles, probably alpha particles, together with two electronic in character, emerging from a point near the surface of the piston (1934).

pected advantage of greatly enhancing the number of associated rays photographed. This was so because the greater the number of rays in a shower of cosmic ray particles, the greater the chance that the counter system controlling the chamber would be set off. As a result the larger showers appeared in the photographs far more frequently relative to single rays than they actually occur in nature. This property of bias towards complex and so interesting phenomena has proved one of the most important advantages of the counter-controlled method.

In a subsequent paper I sketched in detail the formation of tracks by the counter-controlled method and calculated the expected breadth of a track as a function of the coefficient of diffusion of the gaseous ions and of the time elapsing between the passage of the rays and the completion of the expansion. The experimentally measured breadths in hydrogen and oxygen agreed well with the theory.

One serious disadvantage of the counter-controlled method lay in the necessity to maintain the magnetic field for deflecting the particles during the whole period when the apparatus was awaiting the arrival of a ray; in con-

trast when using the random method, the magnetic field could be flashed up momentarily at the moment of the expansion, so avoiding overheating the coils. This demand for a large magnetic field over a large volume, but using only a relatively small expenditure of electric power, led to the design of a special magnet illustrated in Fig. 10. Weighing some 10,000 kg it gave a field up to 14,000 gauss between pole pieces 25 cm in diameter and 15 cm apart, for a power consumption of 25 kW. Cooling was by an air stream from a fan. A shower of high energy taken with the new chamber is shown in Fig. 11, and a single ray of very high momentum traversing a lead plate in Fig. 12.

The first work undertaken with the new apparatus was the measurement of the momentum spectrum of the cosmic ray particles. The earlier measurements by Kunze and by Anderson using the random method had shown the approximate equality of positive and negative momenta up to values of

Fig. 9. Pair of positive and negative electrons produced by gamma rays. (Chadwick, Blackett, and Occhialini, 1934)

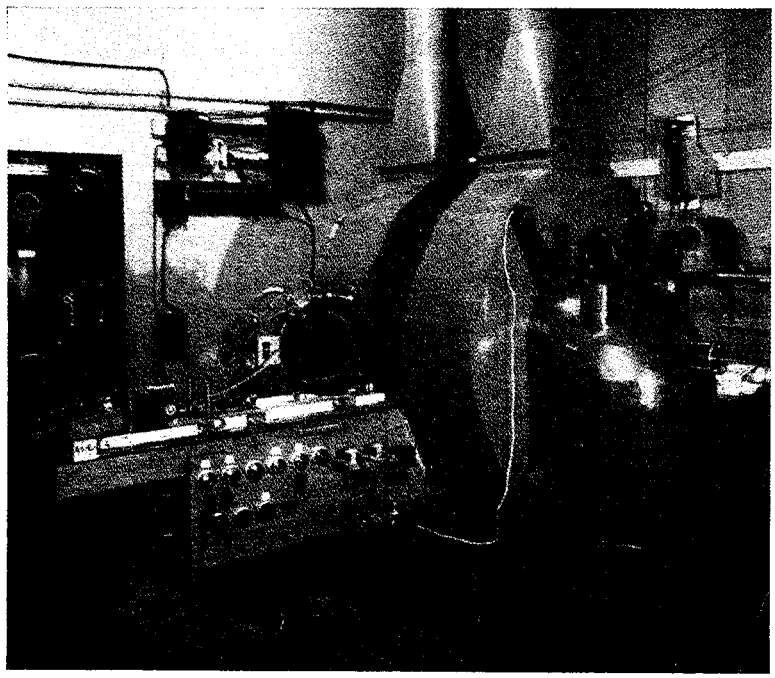


Fig. 10. Magnet and cloud chamber. Ten-ton air-cooled magnet. A 30-cm cloud chamber, slides on rails between the poles of the magnet. On the left of the slide can be seen the automatic release and resetting mechanisms (1935). Figures No. 11 to 19 are all made with this magnet and chamber, though the depth of the chamber and the counter arrangement are different in the later work.

5×10^9 eV/c. With a specially constructed chamber in the field of the new magnet the spectrum was extended in collaboration with R. B. Brode up to about 2×10^{10} eV/c, and it was shown that the differential energy spectrum above 10^9 eV/c could be represented by $g(E)\mathrm{d}E \propto E^{-2}\mathrm{d}E$.

To attain such a high precision of measurement entailed detecting the curvature of a track of 20 cm length when its radius of curvature was 70 metres. Moreover a careful study had to be made of the distortions in the tracks, produced on the one hand by the optical system used, and on the other by the motion of the gas in the interval between the passage of the ray and the instant of the photograph. The great importance of attaining thermal equilibrium in the chamber prior to the expansion came to be recognized.

An optical method of measuring small curvatures was devised by which

the curvature of the image of a reprojected track was compensated by the curvature introduced optically by means of a prism of small angle.

When adequate number of tracks were available it was found that the number of positive tracks was about 13% in excess of the number negatives but the probable error of the determination was rather large, of the order of 6%. Other workers (Hughes and Jones)subsequently found a rather larger positive excess of the order of 25%.

The next task was to measure accurately the loss of energy of the rays in passing through metal plates placed in the chamber, a study initiated by Anderson. This was of special importance in relation to the identification of the penetrating component, which comprised 80% of the rays at sea level and which were far less absorbed than the electronic component. The energy loss of most rays of momentum over 200 MeV/c were found to be quite small, whereas most of the rays of lower momentum were found to be absorbed very rapidly, as was expected from the theory of collision radiation if they were electrons. This result led me to what turned out to be the quite erroneous conclusion that the particles of high momentum were electrons, but with a much smaller energy, less than that given by the quantum theory of radiation. These results were therefore held by me to confirm the views held at that time by many theorists (Nordheim, Williams, and others) that a breakdown of the radiation formulae could be expected at high energies.

Fig. 11. Cosmic ray shower. Nineteen tracks diverging nearly horizontally from a point to one side of the chamber (H = 14,000 gauss). Total energy of visible particles is 5×10^9 eV. Some of the individual particles have an energy of over 10^9 eV (1935).

Subsequent experimental work, particularly by Anderson showed, however, that electrons of high energy did in fact show the large energy loss expected from quantum mechanics, and so proved that the penetrating rays could not be electrons at all, but must be a new type of particle, now called the «meson» and known to have a mass about 200 m. On the theoretical side, Williams and Weiszäcker independently proved by an ingenious application of Fermi's impact parameter method that no breakdown of the radiation formula was to be expected. Final identification of the meson came from the photographs of Street and of Anderson.

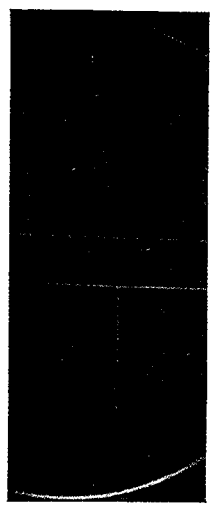


Fig. 12. Mu-meson traversing a 2-cm gold plate. Appreciable energy loss and scattering occurs.

A detailed study in collaboration with J. G. Wilson was made of the scattering of penetrating cosmic rays particles in metal plates and it was shown that the observed scattering agreed closely with that calculated by Williams. A few particles (mainly with a positive charge), were found which showed both abnormally large scattering, and some others which showed abnormally large energy loss. These were thought then to be possibly protons, but the subsequent discovery by Powell and Occhialini using the photographic emulsion technique of the π-meson of mass about 300 m has made other alternative explanations possible.

In parallel with these developments of the counter-controlled cloud chamber as a precision method for measuring the momentum and energy loss of

single cosmic ray particles, a number of investigations were made by many different workers in many different countries of the rarer types of cosmic ray showers, utilizing the selective property of the counter-controlled method to reveal them in numbers far above that of their actual occurrence in nature.

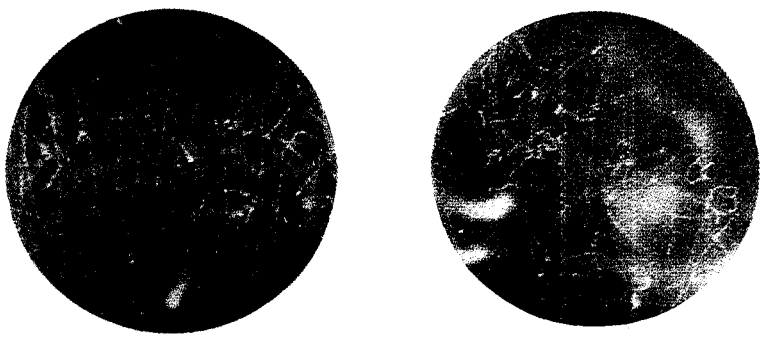


Fig. 13. Extensive shower. Two cloud chambers several metres apart record an extensive air shower. (Wilson and Lovell, 1939)

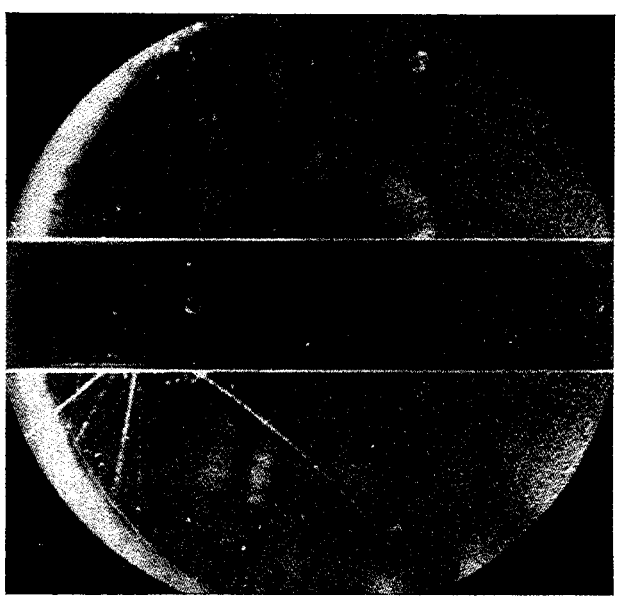


Fig. 14. Explosion showers. Two high-energy incident particles strike a 3.5-cm lead plate: one penetrates without much scattering or energy loss while the other initiates a nuclear explosion. Two heavily ionizing protons are ejected with a few faster particles, which cannot be identified. The markedly curved and heavily ionizing track on the right is identified as a meson. (Rochester, Butler, and Runcorn, 1947)

Particularly beautiful examples of such photographs were made by Anderson, Street, Hazen, Leprince-Ringuet and others and allowed the details of the cascade theory of shower formation to be followed. The extensive air showers, discovered by Auger and his collaborators, were investigated by Lovell and Wilson in my laboratory using two cloud chambers separated by a distance of several metres (Fig. 13). The knock-on showers produced by mesons in lead plates were investigated by Lovell.

In 1939, a counter-controlled cloud chamber was operated by Braddick and Hensby in the Holborn Tube Station in London at a depth of 30 metres underground. Amongst the photographs taken were a few which showed the simultaneous occurrence of two associated penetrating particles.

By delaying the expansion for a fraction of a second after the passage of a ray, the ions diffuse a short distance from their places of formation, so producing a broad track in which the separate droplets condensed on each ion

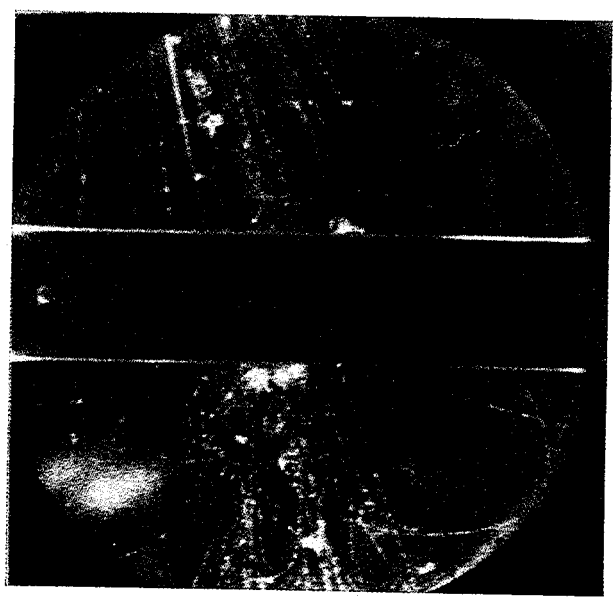


Fig. 15. Penetrating shower with anomalous forked track. A typical, but rare, type of penetrating shower showing several particles penetrating a 3-cm lead plate together with some soft electronic component. On the right below the plate is a peculiar forked track, which for reasons given in the text, is considered to represent the spontaneous disintegration of a new type of neutral particle (τ-meson) of mass about 900 m into a positive and negative particle of lower mass. H = 3,500 gauss. (Rochester and Butler, 1947)

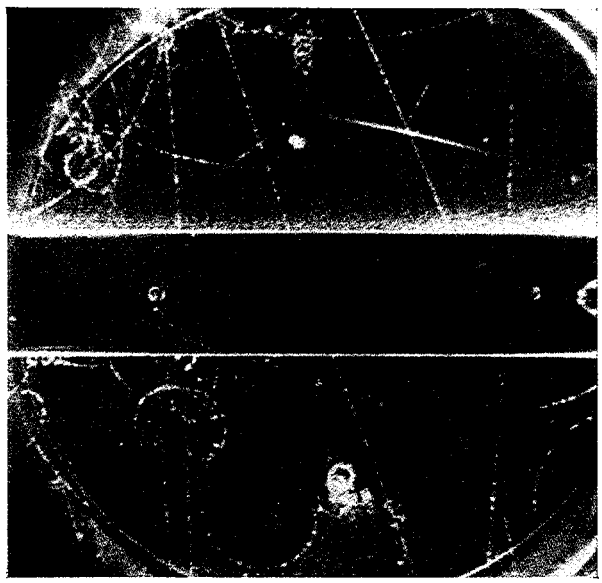


Fig. 16. Penetrating shower with anomalous bent track. A few penetrating particles pass through the plate. One of them, at the top right-hand corner of the photograph, makes an 18° deflection in the gas and then passes through the plate with little further deflection. This is interpreted as the spontaneous disintegration of a new type of positive particle (τ-meson) of mass about 900 m into a positive particle of lower mass together with a neutral unobserved particle. H = 7,000 gauss. (Rochester and Butler, 1947)

can be counted. By this technique it is possible to count the number of ions produced by rays of given momentum. Theoretical considerations by Bethe and by Williams had shown that this ionization should increase slowly with the momentum of the particle when this was much greater than mc. As shown by Williams, this increase arises in a very simple ray from the principle of relativity. Corson and Brode in 1938 in Berkeley succeeded in showing that the predicted increase of ionization does occur with electrons. Sen Gupta in 1940 investigated this phenomenon using the cloud chamber in the big magnet and verified that for electrons the increase of ionization agreed closely with the production up to energies of some 800 MeV at which energy the ionization is some 70% above the minimum.

The counter experiments of Jánossy and Wataghin at sea level showed that a rare type of shower existed consisting of a few associated penetrating rays. These penetrating showers were studied by Rochester during the War with a cloud chamber without a magnetic field and it was shown that some

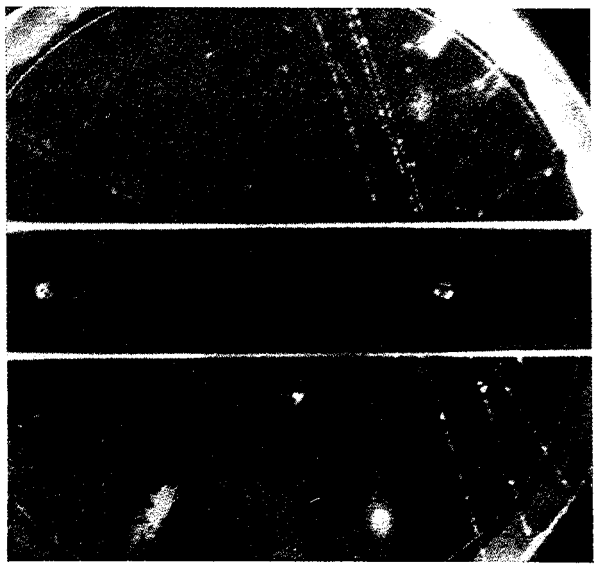


Fig. 17. Penetrating showers with no electronic component. Four associated positive particles traverse a 3.5 cm lead plate. The momenta of three of them are 3.3, 0.9 and 1.0×10^9 eV/c. One is anomalously scattered through 13°. The nature of the particles in such showers is not yet certain, but they may consist of a mixture of protons and positive π-mesons. H = 7,000 gauss (Rochester and Butler, 1947)

of these showers consisted of narrow groups of nearly parallel-penetrating particles. In some experiments, the counter system used to control the chamber was so rigidly defined as to be actuated only once every 24 hours or so. By this means it was possible to set a chamber so that it would wait a whole day to be activated by the particular type of shower which it was desired to photograph.

As soon as the end of the War made it possible to resume work with the large magnet, a detailed study was commenced by Rochester and Butler of the penetrating showers. Jánossy had deduced from his counter experiment that these showers were probably produced by a primary proton component, but the exact nature of the particles in the showers and of the processes by which they are produced remained, and still remain, very obscure. It is generally agreed, however, that the quantum-mechanical treatment of the collision of energetic nucleons with nuclei would be expected to lead to the emission of numbers of ejected protons and mesons. Detailed calculations have been made by Hamilton, Heitler, and Peng, and by Jánossy.

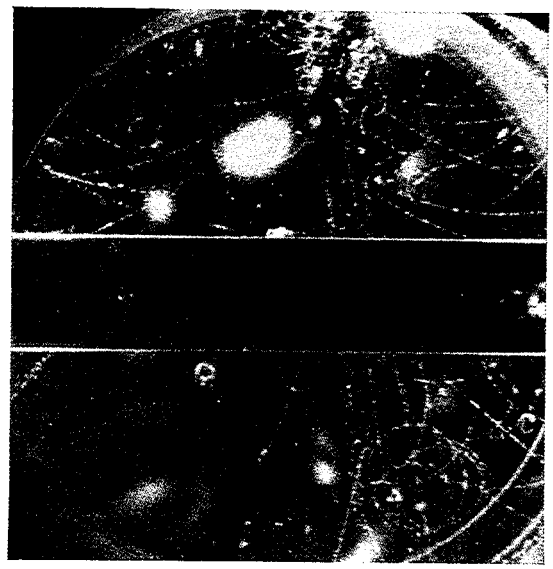


Fig. 18. Penetrating shower with electronic component. Some 8 penetrating particles traverse the plate. Most of them appear to be positive, but this is not certain. A considerable electronic component is present. $H = 7{,}000$ gauss. (Rochester and Butler, 1947)

Among the many thousand photographs taken by Rochester and Butler many interesting phenomena were observed, the details of which are still being elucidated. Of particular importance was the discovery that the large majority of penetrating particles in penetrating showers have a positive charge. Some few of them appear to be protons, but some are certainly not. Since many of these particles appear to be rather highly scattered in a lead plate, it is probable that those that are not protons may be π- or τ-mesons, rather than μ-mesons, which are known to be very little scattered. Certain photographs showed explosive showers in which a number of extremely energetic rays are emitted at rather wide angles, while in others the rays are nearly parallel. It is not yet certain whether these represent two distinct types of showers or whether they represent different aspects of essentially the same phenomenon.

Two photographs taken by Rochester and Butler were of exceptional interest in that they seemed to suggest the existence of two new types of particles, one uncharged and one with a positive charge, and both of mass about 900 m. In one, Fig. 15, a forked track was observed in the gas, due to two particles, one positive and one negative with momenta of a few hundred

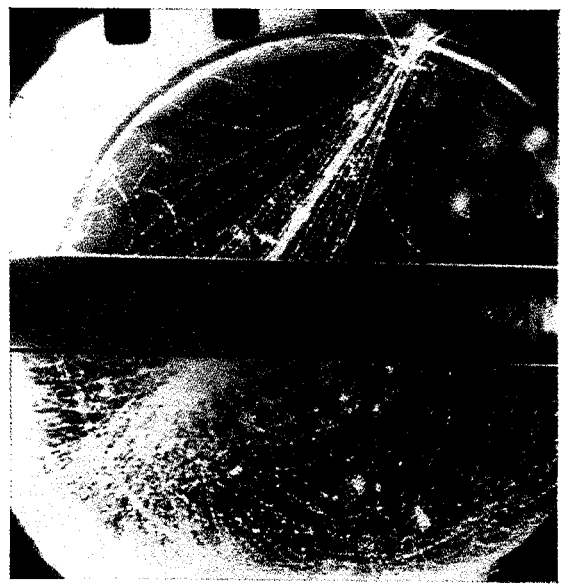


Fig. 19. Cascade shower initiated by meson. This unusual photograph is interpreted as initiated by a meson of very high energy (10^{11} eV/c) emitting a collision radiation, or knocking on an electron, in a lead block above the chamber. Some 40 rays, half positive and half negative, with a total energy of over 10^{10} eV are seen in the top of the chamber and a very large number, perhaps 500 or so, appear below the lead plate. Their number and energies are consistant with a cascade shower of total energy 10^{11} eV, having been initiated a few cascade units back in the lead block over the chamber. A surprising feature is the occurrence of a 9-pronged star of protons, etc. with a total energy of 10^{9} eV originating in the central core of the shower. The explanation of the occurrence is not yet clear, as it would not be expected if the shower is a pure cascade. (Butler, 1948; unpublished.)

MeV/c. The simplest explanation was that a neutral particle had collided with a nucleus and ejected two mesons, but this was rejected since one would expect to find very many more of such cases occurring in the lead plate than in the gas. As these were not found, it was concluded that the forked track did not represent a collision process at all, but a case of spontaneous disintegration of an unstable particle. From the momenta of the ejected particles the mass of the neutral particle was estimated as probably about 870 $\pm$ 200.

A second photograph (Fig. 16) showed a positive particle which seemed to undergo a deflection of 18° in the gas, and then to pass through the 3-cm lead plate without appreciable further deflection or energy loss. Similar argu-

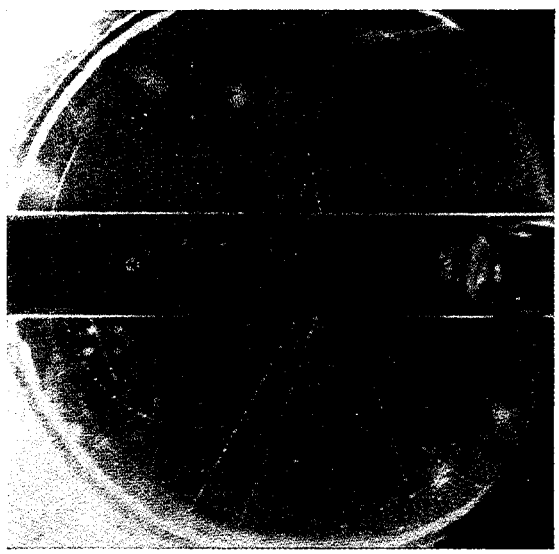


Fig. 20. Wide-angle explosive shower. At least seven particles of momenta about 10^9 eV/c are emitted over a wide solid angle. Two are protons; the rest may be some type of meson. (Butler and Rosser, 1948)

ments to those used for the first photograph led to the interpretation that an unstable positive particle of mass about 1080 ± 100 had spontaneously transformed itself into a positive particle, probably a μ-meson, and into an unobserved neutral particle.

Though an extensive search has been made for further events of this kind, none have been found. However, recent work by Powell in Bristol has given one track which seems must be interpreted as due to a meson of mass about 900. A single track photographed by Leprince-Ringuet in 1940 had been interpreted as indicating a particle of about the same mass. These are now called τ-mesons, and their life is estimated as of the order of 10^{-8} to 10^{-7} seconds.

Although a careful search has been made, no case of a negative proton has yet been found.

Biography

Patrick Maynard Stuart Blackett was born on 18th November, 1897, the son of Arthur Stuart Blackett. He was originally trained as a regular officer for the Navy (Osborne Naval College, 1917; Dartmouth, 1912), and started his career as a naval cadet (1914), taking part, during the First World War, in the battles of Falkland Islands and Jutland. At the end of the war he resigned with the rank of Lieutenant, and took up studies of physics under Lord Rutherford at Cambridge.

After having taken his B.A. degree in 1921, he started research with cloud chambers which resulted, in 1924, in the first photographs of the transmutation of nitrogen into an oxygen isotope. During 1924–1925 he worked at Göttingen with James Franck, after which he returned to Cambridge. In 1932, together with a young Italian scientist, G.P.S. Occhialini, he designed the counter-controlled cloud chamber, a brilliant invention by which they managed to make cosmic rays take their own photographs. By this method the cloud chamber is brought into function only when the impulses from two Geiger-Müller tubes, placed one above and one below the vertical Wilson chamber, coincide as the result of the passing of an electrically charged particle through both of them.

In the spring of 1933 they not only confirmed Anderson's discovery of the positive electron, but also demonstrated the existence of «showers» of positive and negative electrons, both in approximately equal numbers. This fact and the knowledge that positive particles (positrons) do not normally exist as normal constituents of matter on the earth, formed the basis of their conception that gamma rays can transform into two material particles (positrons and electrons), plus a certain amount of kinetic energy – a phenomenon usually called *pair production*. The reverse process – a collision between a positron and an electron in which both are transformed into gamma radiation, so-called *annihilation radiation* – was also verified experimentally. In the interpretation of these experiments Blackett and Occhialini were guided by Dirac's theory of the electron.

Blackett became Professor at Birkback College, London, in 1933, and

there continued cosmic ray research work, hereby collecting a cosmopolitan school of research workers. In 1937 he succeeded Sir Lawrence Bragg at Manchester University, Bragg himself having succeeded Rutherford there; his school of cosmic research work continued to develop, and since the war the Manchester laboratory has extended its field of activity, particularly into that of the radar investigation of meteor trails under Dr. Lovell.

At the start of World War II, Blackett joined the Instrument Section of the Royal Aircraft Establishment. Early in 1940, he became Scientific Advisor to Air Marshall Joubert at Coastal Command, and started the analytical study of the anti U-boat war, building up a strong operational research group. In the same year he became Director of Naval Operational Research at the Admiralty, and continued the study of the anti U-boat war and other naval operations: later in 1940 he was appointed Scientific Advisor to General Pile, C.M.C., Anti-Aircraft Command, and built up an operational research group to study scientifically the various aspects of Staff work. During the blitz he was also concerned with the employment and use of anti-aircraft defence of England.

In 1945, at the end of the Second World War, work was resumed on cosmic ray investigations in the University of Manchester: in particular on the further study of cosmic ray particles by the counter-controlled cloud chamber in a strong magnetic field, built and used before the War. In 1947, Rochester and Butler, working in the laboratory, discovered the first two of what is now known to be a large family of the so-called strange particles. They identified one charged and one uncharged particle which were intrinsically unstable and decayed with a lifetime of some 10^{-10} of a second into lighter particles. This result was confirmed a few years later by Carl Anderson in Pasadena.

Soon after this discovery, the magnet and cloud chamber were moved to the Pic du Midi Observatory in the Pyrenees in order to take advantage of the greater intensity of cosmic ray particles at a very high altitude. This move was rewarded almost immediately by the discovery by Butler and coworkers, within a few hours of starting work, of a new and still stranger strange particle, which was called the negative cascade hyperon. This was a particle of more than protonic mass which decayed into a π-meson and another unstable hyperon, also of more than protonic mass, which itself decayed into a proton and π-meson.

In 1948 Blackett followed up speculations about the isotropy of cosmic rays and began speculating on the origin of the interstellar magnetic fields,

and in so doing revived interest in some 30-year old speculations of Schuster and H. A. Wilson, and others, on the origin of the magnetic field of the earth and sun. Although these speculations are not now considered as likely to be valid, they led him to interest in the history of the earth's magnetic field, and so to the newly born subject of the study of rock magnetism.

Professor Blackett was appointed Head of the Physics Department of the Imperial College of Science and Technology, London, in 1953 and retired in July, 1963. He is continuing at the Imperial College as Professor of Physics and Pro-Rector.

Over the last ten years or so a group under his direction have studied many aspects of the properties of rocks with the object of finding out the precise history of the earth's magnetic field, in magnitude and direction back to the earliest geological times. Such results, together with those of workers in many other countries, seem to indicate that the rock magnetism data supports strongly the conclusions of Wegener and Du Toit that the continents have drifted relative to each other markedly in the course of geological history.

The study is now being continued, directed to explaining the remarkable phenomenon that about 50% of all rocks are reversely magnetized. The experiments are directed towards deciding whether this reversed magnetization is due to reversal of the earth's magnetic field or to a complicated physical or chemical process occurring in the rocks.

Blackett was awarded the Royal Medal by the Royal Society in 1940 and the American Medal for Merit, for operational research work in connection with the U-boat campaign, in 1946. He is the author of *Military and Political Consequences of Atomic Energy* (1948; revised edition 1949; American edition *Fear, War, and the Bomb*, 1949).

In 1924 he married Constanza Bayon; they have one son and one daughter.

Physics 1949

HIDEKI YUKAWA

«for his prediction of the existence of mesons on the basis of theoretical work on nuclear forces»

Physics 1949

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

Royal Hignesses, Ladies and Gentlemen.

For many ages, an important aim of science has been to explain the phenomena we observe by the properties of fundamental particles. In modern physics this problem is of first importance. During the last decade, fundamental particles called « mesons » have turned out to be particularly interesting. The mesons are particles heavier than the electrons but lighter than the nuclei of the hydrogen atoms, i.e. the protons.

The mesons were entirely unknown until Hideki Yukawa in 1934 predicted their existence on the basis of a theoretical investigation of the nuclear forces. This is the achievement which has now been rewarded by the Nobel Prize in Physics.

From earlier research by Heisenberg and others one knew that an atomic nucleus, i.e. the central core of an atom, is composed of protons and of other particles which have the same mass as the protons but no charge. These building elements of the atomic nuclei are called « nucleons » and are held together by the so-called nuclear forces.

Attacking the problem of the nature of the nuclear forces, Yukawa used the electromagnetic field as a model. He found that this field could be modified so as to give forces which like the nuclear forces have a short range. He therefore assumed that the new field corresponds to the nuclear forces. Each field of force is, according to modern theories, associated with some kind of particles. Yukawa discovered that there is a simple relation between the range of the forces and the mass of the corresponding particles. He estimated the range from known experimental data and found that the new particles should be about 200 times heavier than the electrons. The name of mesons for these particles was not introduced until later. According to Yukawa's theory, the nuclear forces can be traced back to an exchange of mesons between the nucleons. These are continually emitting and absorbing mesons.

Yukawa also studied the important question of whether the mesons can appear outside the nuclei. He found that the mesons can be created during the interaction of nucleons if these can deliver a sufficient amount of energy.

Therefore, mesons cannot be created in ordinary nuclear reactions. Yukawa emphasized, however, that they can be expected to appear in the cosmic radiation, in which particles of great energy are found.

Yukawa assumed that mesons can have both positive and negative charge and that the magnitude of the charge is the same as that of the electron. A theory of Fermi, which had been proposed some years earlier, led Yukawa to the assumption, that a meson can be transformed into an electron and a light particle without charge called «neutrino». As was pointed out later, free mesons could therefore be supposed to exist only for a very short time, some millionth of a second or less.

As Yukawa had suggested, the study of the cosmic radiation gave the first experimental evidence of the existence of mesons. This evidence was given in 1937 by Anderson and Neddermeyer and other American physicists. Since that time, the mesons in the cosmic radiation have been very much studied. These investigations have been guided by the theory of Yukawa. A new period in meson research began about three years ago. The British physicist Powell and his collaborators then found that there exist two kinds of mesons. The mesons of one kind are those found in 1937, whereas the mesons of the other kind are somewhat heavier and different also in other respects. Mesons can now be produced in the large cyclotron in Berkeley, California. This has greatly increased the possibilities of studying them.

These experimental investigations have shown, that the masses of both kinds of mesons agree with Yukawa's prediction as far as the order of magnitude is concerned. The heavier mesons, but not the lighter ones, have an interaction with the nucleons about as strong as Yukawa had postulated. The fact that particles of this kind have been found experimentally provides a brilliant vindication of Yukawa's fundamental ideas. The electric charge of both kinds of mesons agree with Yukawa's prediction. It has also been experimentally confirmed, that the mesons can exist only for a very short time. A heavy meson lives only for about one hundredth of a millionth of a second and is then transformed into a light meson and probably a neutrino. The light meson disappears after a few millionths of a second, and electrons are then created and probably also neutrinos.

After experimental evidence of the existence of mesons had been given the interest in Yukawa's theory rose quickly. Much effort was expanded in developing the theory and investigating its consequences. In this work Yukawa and his Japanese collaborators took the lead. Among other things, they found theoretically that neutral mesons exist besides the charged ones.

It has not yet been possible to give a theory for the nuclear forces, which yields results that are in good quantitative agreement with the experiments. Yukawa's theory has, however, led to many important qualitative results about the nuclei. The theory has also proved to be of great value in cosmic-ray research. It was e.g. possible to understand, that mesons can be created in the upper layers of the atmosphere by the primary cosmic radiation falling on the earth.

The research on mesons will probably lead to new discoveries. The meson theory may develop into other forms. By having predicted the existence of the mesons and many of their essential properties Yukawa has accomplished pioneering research of utmost importance. His ideas have proved to be an enormous stimulus to the research in theoretical as well as experimental physics.

Professor Hideki Yukawa. In 1934, when you were only 27 years old, you boldly predicted the existence of new particles, now called « mesons », which you anticipated to be of fundamental importance for the understanding of the forces acting in the atomic nucleus. Recent experiments have provided brilliant support for your essential ideas. These ideas have been exceedingly fruitful and are a guiding star in present-day theoretical and experimental work on atomic nuclei and on cosmic rays. You have also contributed much to other problems in basic theory and you have played a great role in bringing your country to its very high position in modern physical research.

On behalf of the Royal Swedish Academy of Sciences, I wish to congratulate you on your ingenious work, and I now ask you to receive your Nobel Prize from the hands of His Royal Highness the Crown Prince.

HIDEKI YUKAWA

Meson theory in its developments

Nobel Lecture, December 12, 1949

The meson theory started from the extension of the concept of the field of force so as to include the nuclear forces in addition to the gravitational and electromagnetic forces. The necessity of introduction of specific nuclear forces, which could not be reduced to electromagnetic interactions between charged particles, was realized soon after the discovery of the neutron, which was to be bound strongly to the protons and other neutrons in the atomic nucleus. As pointed out by Wigner[1], specific nuclear forces between two nucleons, each of which can be either in the neutron state or the proton state, must have a very short range of the order of 10^{-13} cm, in order to account for the rapid increase of the binding energy from the deuteron to the alpha-particle. The binding energies of nuclei heavier than the alpha-particle do not increase as rapidly as if they were proportional to the square of the mass number A, i.e. the number of nucleons in each nucleus, but they are in fact approximately proportional to A. This indicates that nuclear forces are saturated for some reason. Heisenberg[2] suggested that this could be accounted for, if we assumed a force between a neutron and a proton, for instance, due to the exchange of the electron or, more generally, due to the exchange of the electric charge, as in the case of the chemical bond between a hydrogen atom and a proton. Soon afterwards, Fermi[3] developed a theory of beta-decay based on the hypothesis by Pauli, according to which a neutron, for instance, could decay into a proton, an electron, and a neutrino, which was supposed to be a very penetrating neutral particle with a very small mass.

This gave rise, in turn, to the expectation that nuclear forces could be reduced to the exchange of a pair of an electron and a neutrino between two nucleons, just as electromagnetic forces were regarded as due to the exchange of photons between charged particles. It turned out, however, that the nuclear forces thus obtained was much too small[4], because the beta-decay was a very slow process compared with the supposed rapid exchange of the electric charge responsible for the actual nuclear forces. The idea of the meson field was introduced in 1935 in order to make up this gap[5]. Original assumptions of the meson theory were as follows:

I. The nuclear forces are described by a scalar field U, which satisfies the wave equation

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\varkappa^2\right)U=0 \tag{1}$$

in vacuum, where $\varkappa$ is a constant with the dimension of reciprocal length. Thus, the static potential between two nucleons at a distance r is proportional to $\exp(-\varkappa r)/r$, the range of forces being given by $1/\varkappa$.

II. According to the general principle of quantum theory, the field U is inevitably accompanied by new particles or quanta, which have the mass

$$\mu=\frac{\varkappa\hbar}{c} \tag{2}$$

and the spin 0, obeying Bose-Einstein statistics. The mass of these particles can be inferred from the range of nuclear forces. If we assume, for instance, $\varkappa = 5\times 10^{12}\ \mathrm{cm}^{-1}$, we obtain $\mu \cong 200\, m_e$, where m_e is the mass of the electron.

III. In order to obtain exchange forces, we must assume that these mesons have the electric charge $+e$ or $-e$, and that a positive (negative) meson is emitted (absorbed) when the nucleon jumps from the proton state to the neutron state, whereas a negative (positive) meson is emitted (absorbed) when the nucleon jumps from the neutron to the proton. Thus a neutron and a proton can interact with each other by exchanging mesons just as two charged particles interact by exchanging photons. In fact, we obtain an exchange force of Heisenberg type between the neutron and the proton of the correct magnitude, if we assume that the coupling constant g between the nucleon and the meson field, which has the same dimension as the elementary charge e, is a few times larger than e.

However, the above simple theory was incomplete in various respects. For one thing, the exchange force thus obtained was repulsive for triplet S-state of the deuteron in contradiction to the experiment, and moreover we could not deduce the exchange force of Majorana type, which was necessary in order to account for the saturation of nuclear forces just at the alpha-particle. In order to remove these defects, more general types of meson fields including vector, pseudoscalar and pseudovector fields in addition to the scalar fields, were considered by various authors[6]. In particular, the vector field

was investigated in detail, because it could give a combination of exchange forces of Heisenberg and Majorana types with correct signs and could further account for the anomalous magnetic moments of the neutron and the proton qualitatively. Furthermore, the vector theory predicted the existence of non-central forces between a neutron and a proton, so that the deuteron might have the electric quadripole moment. However, the actual electric quadripole moment turned out to be positive in sign, whereas the vector theory anticipated the sign to be negative. The only meson field, which gives the correct signs both for nuclear forces and for the electric quadripole moment of the deuteron, was the pseudoscalar field[7]. There was, however, another feature of nuclear forces, which was to be accounted for as a consequence of the meson theory. Namely, the results of experiments on the scattering of protons by protons indicated that the type and magnitude of interaction between two protons were, at least approximately, the same as those between a neutron and a proton, apart from the Coulomb force. Now the interaction between two protons or two neutrons was obtained only if we took into account the terms proportional to g^4, whereas that between a neutron and a proton was proportional to g^2, as long as we were considering charged mesons alone. Thus it seemed necessary to assume further:

IV. In addition to charged mesons, there are neutral mesons with the mass either exactly or approximately equal to that of charged mesons. They must also have the integer spin, obey Bose-Einstein statistics and interact with nucleons as strongly as charged mesons.

This assumption obviously increased the number of arbitrary constants in meson theory, which could be so adjusted as to agree with a variety of experimental facts. These experimental facts could not be restricted to those of nuclear physics in the narrow sense, but was to include those related to cosmic rays, because we expected that mesons could be created and annihilated due to the interaction of cosmic ray particles with energies much larger than μc^2 with matter. In fact, the discovery of particles of intermediate mass in cosmic rays in 1937[8] was a great encouragement to further developments of meson theory. At that time, we came naturally to the conclusion that the mesons which constituted the main part of the hard component of cosmic rays at sea level was to be identified with the mesons which were responsible for nuclear forces[9]. Indeed, cosmic ray mesons had the mass around 200 m_e as predicted and moreover, there was the definite evidence for the spontaneous decay, which was the consequence of the following assumption of the original meson theory:

V. Mesons interact also with light particles, i.e. electrons and neutrinos, just as they interact with nucleons, the only difference being the smallness of the coupling constant g' in this case compared with g. Thus a positive (negative) meson can change spontaneously into a positive (negative) electron and a neutrino, as pointed out first by Bhabha[10]. The proper lifetime, i.e. the mean lifetime at rest, of the charged scalar meson, for example, is given by

$$\tau_0 = 2\left(\frac{\hbar c}{(g')^2}\right)\left(\frac{\hbar}{\mu c^2}\right) \tag{3}$$

For the meson moving with velocity v, the lifetime increases by a factor $1/\sqrt{1-(v/c)^2}$ due to the well-known relativistic delay of the moving clock. Although the spontaneous decay and the velocity dependence of the lifetime of cosmic ray mesons were remarkably confirmed by various experiments[11], there was an undeniable discrepancy between theoretical and experimental values for the lifetime. The original intention of meson theory was to account for the beta-decay by combining the assumptions III and V together. However, the coupling constant g', which was so adjusted as to give the correct result for the beta-decay, turned out to be too large in that it gave the lifetime τ_0 of mesons of the order of 10^{-8} sec, which was much smaller than the observed lifetime 2×10^{-6} sec. Moreover, there were indications, which were by no means in favour of the expectation that cosmic-ray mesons interacted strongly with nucleons. For example, the observed cross-section of scattering of cosmic-ray mesons by nuclei was much smaller than that obtained theoretically. Thus, already in 1941, the identification of the cosmic-ray meson with the meson, which was supposed to be responsible for nuclear forces, became doubtful. In fact, Tanikawa and Sakata[12] proposed in 1942 a new hypothesis as follows: The mesons which constitute the hard component of cosmic rays at sea level are not directly connected with nuclear forces, but are produced by the decay of heavier mesons which interacted strongly with nucleons.

However, we had to wait for a few years before this two-meson hypothesis was confirmed, until 1947, when two very important facts were discovered. First, it was discovered by Italian physicists[13] that the negative mesons in cosmic rays, which were captured by lighter atoms, did not disappear instantly, but very often decayed into electrons in a mean time interval of the order of 10^{-6} sec. This could be understood only if we supposed that ordinary mesons in cosmic rays interacted very weakly with nucleons. Soon after-

wards, Powell and others[14] discovered two types of mesons in cosmic rays, the heavier mesons decaying in a very short time into lighter mesons. Just before the latter discovery, the two-meson hypothesis was proposed by Marshak and Bethe[15] independent of the Japanese physicists above mentioned. In 1948, mesons were created artificially in Berkeley[16] and subsequent experiments confirmed the general picture of two-meson theory. The fundamental assumptions are now[17]:

(i) The heavier mesons, i.e. π-mesons with the mass m_π about 280 m_e interact strongly with nucleons and can decay into lighter mesons, i.e. μ-mesons and neutrinos with a lifetime of the order of 10^{-8} sec; π-mesons have integer spin (very probably spin 0) and obey Bose-Einstein statistics. They are responsible for, at least, a part of nuclear forces. In fact, the shape of nuclear potential at a distance of the order of $\hbar/m_\pi c$ or larger could be accounted for as due to the exchange of π-mesons between nucleons.

(ii) The lighter mesons, i.e. μ-mesons with the mass about 210 m_e are the main constituent of the hard component of cosmic rays at sea level and can decay into electrons and neutrinos with the lifetime 2×10^{-6} sec. They have very probably spin $\frac{1}{2}$ and obey Fermi-Dirac statistics. As they interact only weakly with nucleons, they have nothing to do with nuclear forces.

Now, if we accept the view that π-mesons are the mesons that have been anticipated from the beginning, then we may expect the existence of neutral π-mesons in addition to charged π-mesons. Such neutral mesons, which have integer spin and interact as strongly as charged mesons with nucleons, must be very unstable, because each of them can decay into two or three photons[18]. In particular, a neutral meson with spin 0 can decay into two photons and the lifetime is of the order of 10^{-14} sec or even less than that. Very recently, it became clear that some of the experimental results obtained in Berkeley could be accounted for consistently by considering that, in addition to charged π-mesons, neutral π-mesons with the mass approximately equal to that of charged π-mesons were created by collisions of high-energy protons with atomic nuclei and that each of these neutral mesons decayed into two mesons with the lifetime of the order of 10^{-13} sec or less[19]. Thus, the neutral mesons must have spin 0.

In this way, meson theory has changed a great deal during these fifteen years. Nevertheless, there remain still many questions unanswered. Among other things, we know very little about mesons heavier than π-mesons. We do not know yet whether some of the heavier mesons are responsible for nuclear forces at very short distances. The present form of meson theory is

not free from the divergence difficulties, although recent development of relativistic field theory has succeeded in removing some of them. We do not yet know whether the remaining divergence difficulties are due to our ignorance of the structure of elementary particles themselves[20]. We shall probably have to go through another change of the theory, before we shall be able to arrive at the complete understanding of the nuclear structure and of various phenomena, which will occur in high energy regions.

1. E. Wigner, *Phys. Rev.*, 43 (1933) 252.
2. W. Heisenberg, *Z. Physik*, 77 (1932) 1; 78 (1932) 156; 80 (1933) 587.
3. E. Fermi, *Z. Physik*, 88 (1934) 161.
4. I. Tamm, *Nature*, 133 (1934) 981; D. Ivanenko, *Nature*, 133 (1934) 981.
5. H. Yukawa, *Proc. Phys.-Math. Soc. Japan*, 17 (1935) 48; H. Yukawa and S. Sakata, *ibid.*, 19 (1937) 1084.
6. N. Kemmer, *Proc. Roy. Soc. London*, A 166 (1938) 127; H. Fröhlich, W. Heitler, and N. Kemmer, *ibid.*, 166 (1938) 154; H. J. Bhabha, *ibid.*, 166 (1938) 501; E. C. G. Stueckelberg, *Helv. Phys. Acta*, 11 (1938) 299; H. Yukawa, S. Sakata, and M. Taketani, *Proc. Phys.-Math. Soc. Japan*, 20 (1938) 319; H. Yukawa, S. Sakata, M. Kobayasi, and M. Taketani, *ibid.*, 20 (1938) 720.
7. W. Rarita and J. Schwinger, *Phys. Rev.*, 59 (1941) 436, 556.
8. C. D. Anderson and S. H. Neddermeyer, *Phys. Rev.*, 51 (1937) 884; J. C. Street and E. C. Stevenson, *ibid.*, 51 (1937) 1005; Y. Nishina, M. Takeuchi, and T. Ichimiya, *ibid.*, 52 (1937) 1193.
9. H. Yukawa, *Proc. Phys.-Math. Soc. Japan*, 19 (1937) 712; J. R. Oppenheimer and R. Serber, *Phys. Rev.*, 51 (1937) 1113; E. C. G. Stueckelberg, *ibid.*, 52 (1937) 41.
10. H. J. Bhabha, *Nature*, 141 (1938) 117.
11. H. Euler and W. Heisenberg, *Ergeb. Exakt. Naturw.*, 17 (1938) 1; P. M. S. Blackett, *Nature*, 142 (1938) 992; B. Rossi, *Nature*, 142 (1938) 993; P. Ehrenfest, Jr. and A. Fréon, *Compt. Rend.*, 207 (1938) 853; E. J. Williams and G. E. Roberts, *Nature*, 145 (1940) 102.
12. Y. Tanikawa, *Progr. Theoret. Phys. Kyoto*, 2 (1947) 220; S. Sakata and K. Inouye, *ibid.*, 1 (1946) 143.
13. M. Conversi, E. Pancini, and O. Piccioni, *Phys. Rev.*, 71 (1947) 209.
14. C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell, *Nature*, 159 (1947) 694; C. M. G. Lattes, G. P. S. Occhialini, and C. F. Powell, *Nature*, 160 (1947) 453, 486.
15. R. E. Marshak and H. A. Bethe, *Phys. Rev.*, 72 (1947) 506.
16. E. Gardner and C. M. G. Lattes, *Science*, 107 (1948) 270; W. H. Barkas, E. Gardner, and C. M. G. Lattes, *Phys. Rev.*, 74 (1948) 1558.
17. As for further details, *see* H. Yukawa, *Rev. Mod. Phys.*, 21 (1949) 474.

18. S. Sakata and Y. Tanikawa, *Phys. Rev.*, 57 (1940) 548; R. J. Finkelstein, *ibid.*, 72 (1947) 415.
19. H. F. York, B. J. Moyer, and R. Bjorklund, *Phys. Rev.*, 76 (1949) 187.
20. H. Yukawa, *Phys. Rev.*, 77 (1950) 219.

Biography

Hideki Yukawa was born in Tokyo, Japan, on 23rd January, 1907, the third son of Takuji Ogawa, who later became Professor of Geology at Kyoto University. The future Laureate was brought up in Kyoto and graduated from the local university in 1929. Since that time he has been engaged on investigations in theoretical physics, particularly in the theory of elementary particles.

Between 1932 and 1939 he was a lecturer at the Kyoto University and lecturer and Assistant Professor at the Osaka University. Yukawa gained the D.Sc. degree in 1938 and from the following year he has been, and still is, Professor of Theoretical Physics at Kyoto University. While at Osaka University, in 1935, he published a paper entitled « On the Interaction of Elementary Particles. I. » (*Proc. Phys.-Math. Soc. Japan*, 17, p. 48), in which he proposed a new field theory of nuclear forces and predicted the existence of the meson. Encouraged by the discovery by American physicists of one type of meson in cosmic rays, in 1937, he devoted himself to the development of the meson theory, on the basis of his original idea. Since 1947 he has been working mainly on the general theory of elementary particles in connection with the concept of the « non-local » field.

Yukawa was invited as Visiting Professor to the Institute for Advanced Study at Princeton, U.S.A., in 1948, and since July, 1949 he has been Visiting Professor at Columbia University, New York City.

The learned societies of his native land have recognised his ability and he is a member of the Japan Academy, the Physical Society and the Science Council of Japan, and is Emeritus Professor of Osaka University. As Director of the Research Institute for Fundamental Physics in Kyoto University he has his office in the Yukawa Hall, which is named after him. He is also a Foreign Associate of the American National Academy of Sciences and a Fellow of the American Physical Society.

The Imperial Prize of the Japan Academy was awarded to Yukawa in 1940; he received the Decoration of Cultural Merit in 1943, and the crowning award, the Nobel Prize for Physics, in 1949.

A large number of scientific papers have been published by him and many books, including *Introduction to Quantum Mechanics* (1946) and *Introduction to the Theory of Elementary Particles* (1948), both in Japanese, have come from his pen. He has edited a journal in English, *Progress of Theoretical Physics*, since 1946.

An honorary doctorate of the University of Paris and honorary memberships of the Royal Society of Edinburgh, the Indian Academy of Sciences, the International Academy of Philosophy and Sciences, and the Pontificia Academia Scientiarum have marked the recognition he has earned in world scientific circles.

A civic honour was awarded to him when he was created Honorary Citizen of the City of Kyoto, Japan.

In 1932 he married, and he and his wife Sumiko have two sons, Harumi and Takaaki.

Physics 1950

CECIL FRANK POWELL

«for his development of the photographic method of studying nuclear processes and his discoveries regarding mesons made with this method»

Physics 1950

Presentation Speech by Professor A. E. Lindh, member of the Nobel Committee for Physics

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

In awarding this year's Nobel Prize in Physics to Professor C. F. Powell of Bristol, the Swedish Academy of Sciences cited his development of the photographic method for the study of nuclear processes and his discoveries concerning the mesons.

The photographic method used by Professor Powell is based on the fact that after an electrically charged particle has passed through a photographic emulsion, the silver bromide grains of the emulsion can be developed, making the path of the particle appear as a dark line which is, actually, a series of blackened grains with longer or shorter intervals between. The distance between the grains is proportional to the speed of the particle; the greater the speed of the particle, the greater the distance, which circumstance is connected with the fact that a swift particle has less power of ionizing than a slow one.

The method is not new; it came into use in the early years of the 20th century as a means of demonstrating radioactive radiation. For the use of the method in the study of nuclear processes it was first necessary to have emulsions sensitive to various kinds of charged particles, and especially to very swift particles. The problem was brought nearer its solution in the early thirties when it was found that sensitizing the plates made them react to swift protons. The method was difficult, however, and it was not widely used.

Emulsions which reacted to swift protons without previous sensitizing were produced independently in 1935 by Zhdanov in Leningrad and by the Ilford Laboratories.

In nuclear physics the photographic method had not been generally accepted even by the end of the thirties, despite the fact that various researchers had used it for studying cosmic radiation. Nuclear physicists were sceptical of the method because divergent results had been obtained in calculating the energy of the particles from measurements of the length of their traces. They placed more confidence in the so-called « Wilson chamber », where the radia-

tion falls into an expansion chamber filled with moisture-saturated air or another gas. The gas is cooled by suddenly expanding the chamber, and drops of mist are deposited on the ions formed in the path of the particles. Under proper lighting, the paths of the particles which are in the chamber at the moment of expansion appear as cloud tracks.

It is Professor Powell's merit to have dispersed this scepticism regarding the photographic method and to have made it an extremely effective aid in investigating certain nuclear processes as well as cosmic radiation and the resultant nuclear phenomena. Using the new Ilford half-tone plates, he began to investigate the usefulness and reliability of the photographic method for the study of nuclear processes, and in a series of projects from 1939 to 1945 he and his collaborators went into various nuclear processes while introducing successive improvements in the treatment of the material, the research technique, and the optical equipment for analyzing the particle traces. These studies have given convincing proof that in this kind of research the photographic method is quite the equal of the Wilson chamber and counter, and even, in some instances, its superior. The savings of time and material effected with the photographic method have been proved by comparisons between similar investigations made with both the Wilson chamber and the photographic method. In one such experiment with the Wilson chamber 20,000 stereoscopic photos produced 1,600 particle traces suitable for measurement. Professor Powell and his collaborators used 3,000 particle traces found on a three-centimetre square of the photographic plate. An important step forward in their efforts to improve and develop the photographic method was taken in 1946, when Professor Powell and his collaborators reported on experiments with a new Ilford emulsion, called « C 2 », whose properties excelled those of the half-tone emulsion in every respect. The traces of the particles appeared more clearly and the absence of disturbing background considerably increased the reliability of the measurements. It now became possible to make another attempt at solving the problem of discovering rare processes, and to charge the emulsions with other atoms for special investigations. The improved photographic method had the greatest importance for the study of cosmic radiation. When one considers that the photographic plates register continuously, whereas the Wilson chamber, as it were, discovers particles and processes only during the brief moments of exposure, it is easy to see that the photographic method offers great advantages over the Wilson chamber for these investigations. Plates with the new emulsion were exposed to cosmic radiation on Pic du Midi, 2,800 metres

above sea level. During the study of these plates and of plates exposed at higher altitudes, up to 5,500 metres, a great number of isolated particle traces were found, as well as so-called « disintegration stars » with varying numbers of ramifications, originating from the disintegration of atomic nuclei in the emulsion. Analysis of these stars showed that some of them had been produced by a particle of small mass which had entered the emulsion, passed into an atomic nucleus, and caused its disintegration. A more detailed investigation showed that the active particle was a meson, which has a mass a few hundred times greater than that of the electron, and which was, in this case, negatively charged. Some cases of nuclear disintegration were observed in which slow mesons were ejected from the nucleus. Continued investigation of the plate material revealed other remarkable phenomena. In 1947, Powell Occhialini, Muirhead, and Lattes reported the discovery of mesons which at the end of their path give rise to secondary mesons. The analysis of the traces of primary and secondary mesons indicated the probable existence of two kinds of mesons having different masses, a theory which was vindicated by further experiments. The primary mesons were named π-mesons and the secondary mesons, μ-mesons. Preliminary determinations of the masses showed that the mass of the π-meson was greater than that of the μ-meson. The charge was equal to the electrical elementary charge. It would be impossible on this occasion to go into greater detail concerning the ingenious methods invented by the Bristol researchers for identifying the paths of the particles, or about the extensive work which was done to determine the relation between the masses of the π-mesons and μ-mesons, and to investigate their properties. I shall only review very briefly some of the most important conclusions regarding the mesons and their properties.

It was found in Professor Powell's laboratory that the mass of the π-meson was 1.35 times greater than that of the μ-meson, a relation which agreed closely with the value of 1.33 which the Berkeley researchers had determined for artificial mesons produced in their 184-inch cyclotron. The mass of the π-meson is said to be 286 times greater than the mass of the electron, and that of the μ-meson 216 times greater. The latter meson is identical with the one whose existence in cosmic rays had been previously established by American researchers. Both the π- and the μ-meson may be positively or negatively charged. The lifetime of the μ-meson has been found to be one millionth of a second, that of the π-meson one hundred times shorter. The π-mesons are unstable and disintegrate spontaneously into μ-mesons. The negative π-mesons easily enter into reciprocal action with the constituents

of the atomic nucleus and at the end of their paths in the emulsion they are caught by atoms and give rise to the disintegration of both light and heavy atomic nuclei. Thanks to the introduction of a new emulsion which, unlike those mentioned above, is sensitive to electrons (the Kodak N.T.4 emulsion), Professor Powell showed in 1949 that the μ-mesons disintegrate at the end of their paths into one charged light particle and, probably, into at least two neutral particles.

Among Professor Powell's latest investigations I should mention his study of the τ-mesons, which have a mass about 1,000 times that of the electron. Their existence had been established earlier by various researchers, and more proof was gathered through the work at Bristol.

The introduction of the new electron-sensitive emulsions makes it probable that we may expect further important discoveries from Professor Powell's laboratory. One has been reported as recently as this year: the discovery of the neutral meson in cosmic radiation. The existence of this particle had been established earlier in the work with artificial mesons at Berkeley. Its lifetime has been found to be 100 million times shorter than the lifetime of the μ-meson, which is one-millionth of a second.

Professor Powell. One of the many who have proposed you to the Nobel Committee as a candidate for the Physics Prize said, «His special claim to consideration is, in my view, the fact that he has shown that discoveries of fundamental importance can still be made with the simplest apparatus – in this case special nuclear emulsions developed under his general direction – and microscopes.» No one can dispute these facts. Through many years of purposeful work you have brought the photographic method to undreamt-of perfection and have made it one of the most efficient aids of modern nuclear physics. The great variety of investigations into atomic processes which have been conducted with the photographic method at your laboratory have made it abundantly clear that after the introduction of your improvements this method has an uncontested position among the most important tools available to the nuclear physicist of our times. The great superiority and efficiency of your improved method as compared to other methods for the study of cosmic rays have been eloquently and convincingly confirmed by the sensational and significant discoveries made by yourself and your distinguished staff in regard to these rays and the nuclear processes caused by them. Your study of the mesons and your discoveries in this connection have borne new members to the family of elementary particles. I need not stress

the extraordinary importance of your discoveries for research in nuclear physics, more particularly for our concept of nuclear energy and our knowledge of cosmic radiation. I only wish to give expression to the sincere admiration and respect we physicists feel for your eminent work through which, in pursuance of great British traditions, you have enriched our field of knowledge with results of the greatest scientific value.

On behalf of the Royal Swedish Academy of Sciences I wish to congratulate you on your significant work and discoveries and to request that you receive your well-earned reward, the Nobel Prize in Physics for the year 1950, from the hands of His Majesty the King.

CECIL F. POWELL

The cosmic radiation*

Nobel Lecture, December 11, 1950

Coming out of space and incident on the high atmosphere, there is a thin rain of charged particles known as the primary cosmic radiation. As a result of investigations extending over more than 30 years, we now know the nature of the incoming particles, and some, at least, of the most important physical processes which occur as a result of their passage through the atmosphere.

Today the study of the cosmic radiation is, in essence, the study of nuclear physics of the extreme high-energy region. Although the number of incoming particles is very small in comparison with those which are produced by the great machines, most of them are much more energetic than any which we can yet generate artificially; and in nuclear collisions they produce effects which cannot be simulated in the laboratory. The study of the resulting transmutations is therefore complementary to that which can be made at lower energies with the aid of the cyclotrons and synchrotrons.

For the investigation of the cosmic radiation, it is necessary to solve two principal technical problems: First, to detect the radiation, to determine the masses, energy and transformation properties of the particles of which it is composed, and to study the nuclear transmutations which they produce. Second, to develop methods of making such observations throughout the atmosphere and at depths underground.

For the detection of the radiations, the same devices are available as in the general field of nuclear physics, and two main classes can be distinguished.

In the first class are found the trigger mechanisms such as the Geiger counter and the scintillation counter. Such devices record the instants of passage of individual particles through the apparatus. Their most important advantages are (a) that they allow observations to be made of great statistical weight; and (b) that the relationship in time of the instants of passage of associated particles can be established. With modern instruments of this type, the time interval between the arrival of two charged particles can be

* The lecture was illustrated by lantern slides and a film of the construction and launching of balloons.

measured even although this is as small as one or two hundredths of a microsecond. These devices have made possible contributions of the greatest importance to our knowledge of the subject, and they have proved especially valuable when the nature of the physical processes being studied has been well understood.

In the second class of detectors are the devices for making manifest the tracks of particles; namely, the Wilson expansion chamber and the photographic plate. These instruments have the particular advantage, amongst others, that they allow a direct and detailed insight into the physical processes which accompany the passage of charged particles through matter. On the other hand, it is arduous to employ them to obtain observations of great statistical weight. The two classes of instruments thus provide complementary information, and each has made a decisive contribution.

The second principal technical problem to be solved is that of making experiments at great altitudes. Some information has been obtained by means of V-2 rockets which pass almost completely out of the earth's atmosphere, but their time of flight is restricted to only a few minutes. Alternatively, balloons can be made to ascend to great altitudes and to give level flights for many hours. The simplicity of the photographic method of recording the tracks of charged particles makes it very suitable as a detector in such experiments.

Today the most suitable types of balloons for experiments on the cosmic radiation are those made of thin sheets of a plastic material, « polyethylene ». Although rubber balloons can sometimes be made to ascend higher into the atmosphere, their performance is erratic. The rubber, whether natural or synthetic, appears to perish rapidly under the action of the solar radiation high above the clouds: it is therefore difficult, even when employing many rubber balloons in a single experiment, to secure the sustained level flight which is desirable. On the other hand, polyethylene is chemically inert, and the fabric of the balloon can remain for many hours at high altitudes without any serious effect on its mechanical strength.

In Bristol, we construct balloons of polyethylene by methods similar in principle to those developed in the U.S.A. by the General Mills Corporation. We employ polyethylene sheet $1\frac{1}{2}$ thousandths of an inch thick, the shaped pieces of which are « heat-sealed » together to form an envelope which, when fully inflated, is nearly spherical in form. Unlike those of rubber, these balloons are open at the lower end; and just before launching, the envelope is very slack and contains only a small fraction of its total volume filled with

hydrogen (see Fig. 1). As the balloon ascends, the pressure falls and the balloon inflates. Near maximum altitude its envelope becomes tensed and hydrogen escapes from the bottom aperture. Balloons of this type, 20 m in diameter, give – with light loads of about 20 kg – level flights at altitudes of the order of 95,000 ft. It is anticipated that a similar balloon 50 m in diameter should reach about 120,000 ft.

By observations at great altitudes we now know that the primary cosmic radiation is made up of atomic nuclei moving at speeds closely approaching that of light (Freier *et al.*[1]; Bradt and Peters[2]). It is possible to record the tracks of the incoming particles (see Fig. 2) and to determine their charge; and thence the relative abundance of the different chemical elements. Recent experiments prove that hydrogen and helium occur most frequently, and the distribution in mass of the heavier nuclei appears to be similar to that of the matter of the universe. Thus elements more massive than iron or nickel occur, if at all, very infrequently.

The detailed study of the « mass spectrum » of the incoming nuclei has an important bearing on the problem of the origin of the primary particles; but it is complicated by the fact that, because of their large charge, the particles rapidly lose energy in the atmosphere by making atomic and nuclear collisions. They therefore rarely penetrate to altitudes less than 70,000 ft. It is for this reason that exposures at high altitudes are of particular interest.

A second reason for making experiments at extreme altitudes is that the primary nuclei commonly suffer fragmentation in making nuclear collisions (see Fig. 3). A primary nucleus of magnesium or aluminium, for example, may decompose into lighter nuclei such as lithium, α-particles and protons. The mass spectrum at a given depth is therefore different from that of the primary radiation, and such effects are appreciable at 90,000 ft., where the mass of overlying air is about 20 g per cm^2. They would be much reduced at 120,000 ft. (6 g per cm^2), an altitude which, we have seen, appears to be accessible with very large balloons.

The primary protons and α-particles, because of their smaller charge, penetrate to much lower altitudes. In collisions they disintegrate the nuclei which they strike (see Fig. 4) and, in the process, lead to the creation of new forms of matter, the π-mesons of mass 274 m_e (Lattes *et al.*[3]; Piccioni[4]; Fowler[5]). These particles are usually ejected with great speed and proceed downwards towards the earth.

The π-mesons are now known to have an average lifetime of about 2×10^{-8} sec (Richardson[6]). This period is so short that, when moving in a

Fig. 1. Inflation of balloon of polyethylene just after dawn. The balloon has a total length of about 120 ft. and most of the fabric is on the ground. Such a balloon can in favourable conditions give level flight at about 90,000 ft. for many hours with a load of 40 kg.

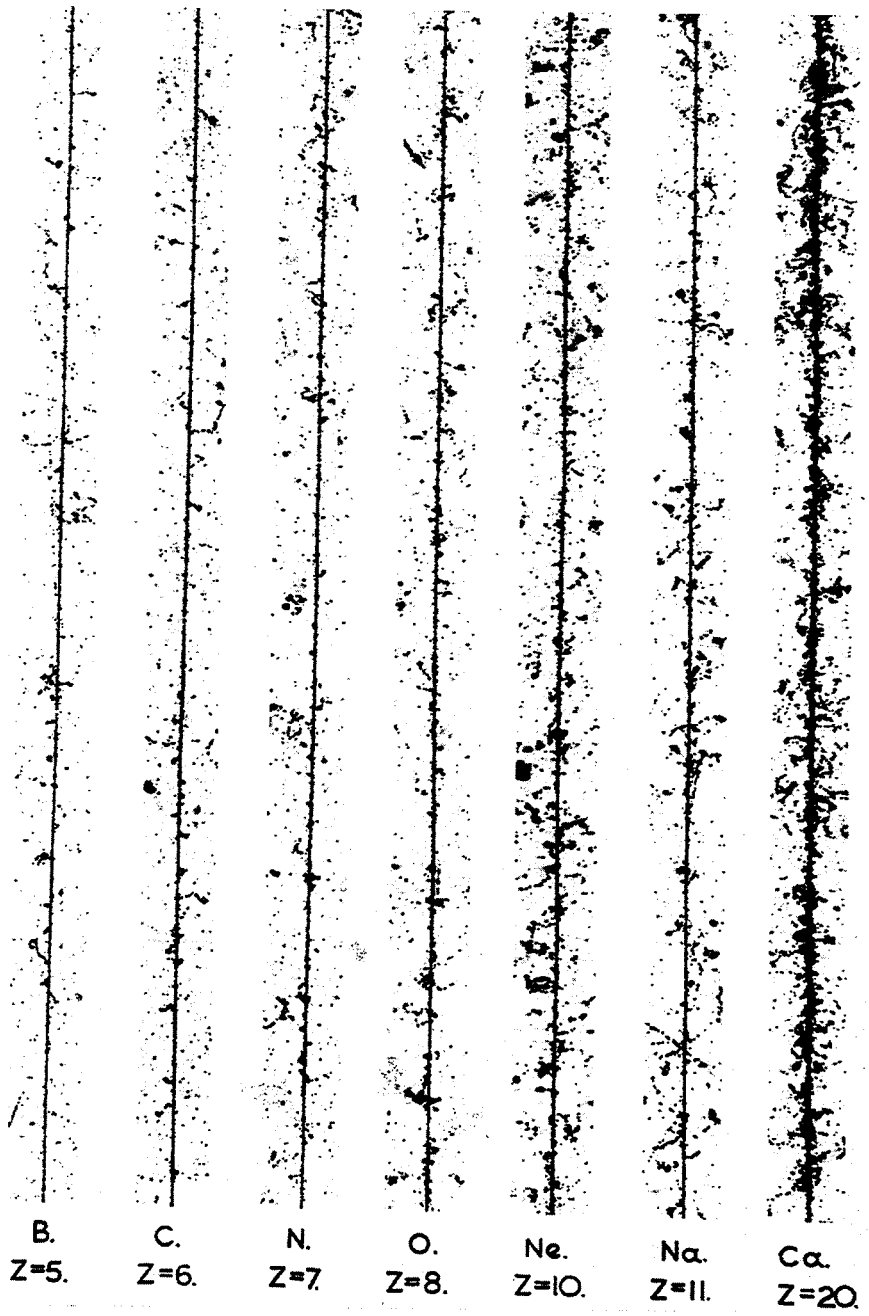


Fig. 2. Examples of the tracks in photographic emulsions of primary nuclei of the cosmic radiation moving at relativistic velocities.

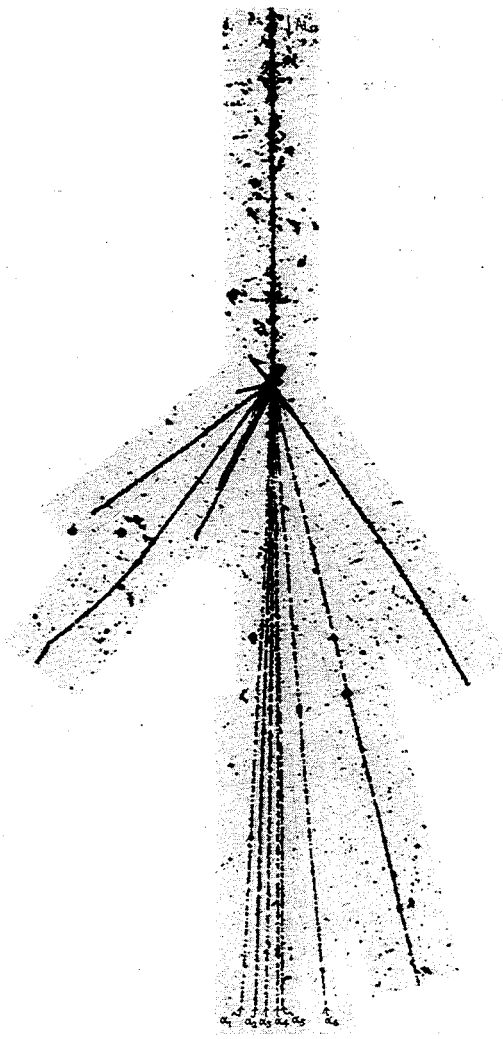


Fig. 3. A nucleus of magnesium or aluminium moving with great velocity, collides with another nucleus in a photographic emulsion. The incident nucleus splits up into six α-particles of the same speed and the struck nucleus is shattered.

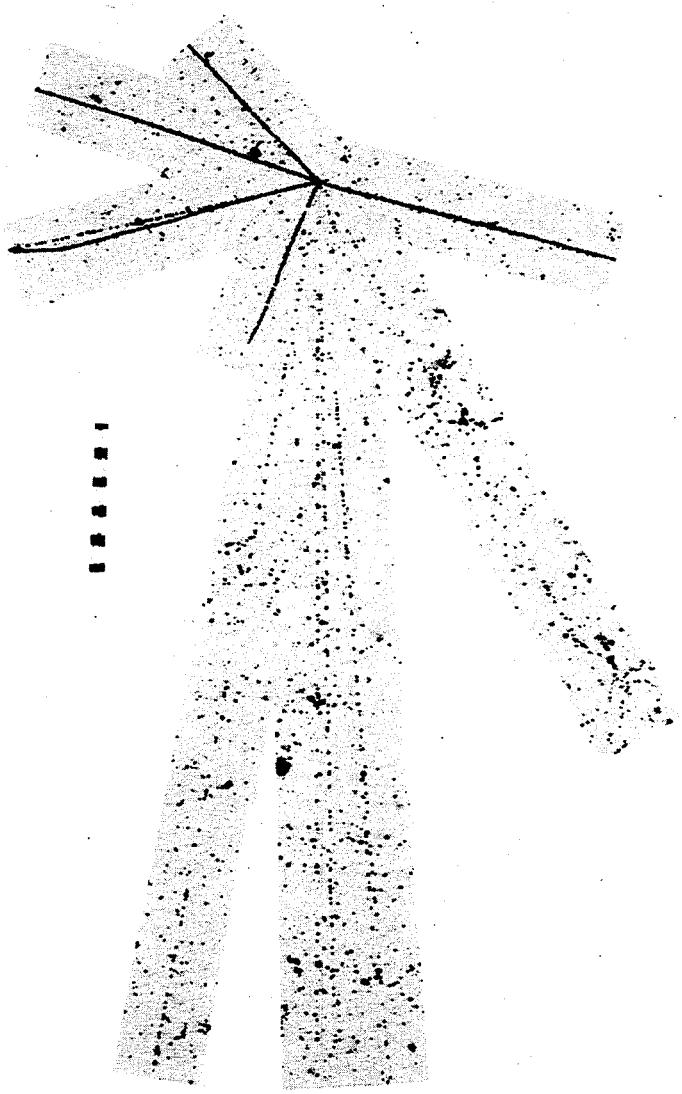


Fig. 4. A disintegration caused by a fast neutron of the cosmic radiation. The «thin tracks» of the fast-moving shower particles can be distinguished. Most of these are due to π-particles. A γ-ray, produced by the decay of a neutral π-meson, gives rise to a pair of electrons near the «star».

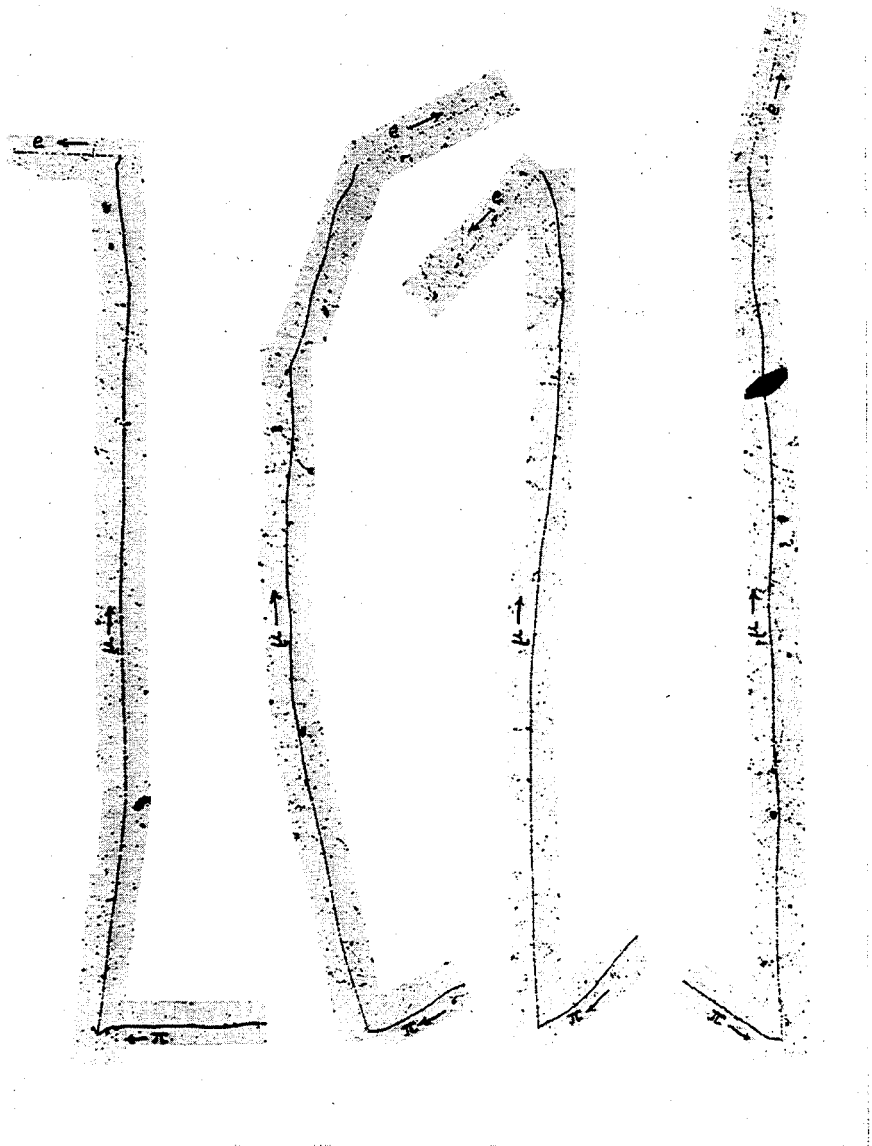


Fig. 5. Photo-micrographs of four examples of the successive decay π-μ-e as recorded in photographic emulsions.

gas – in which their velocity is reduced, by loss of energy through ionization, at a relatively slow rate – they commonly decay in flight. In a solid material, however, they can be arrested before they have had time to decay. This was the most important of the factors which prevented the identification of the particles until after the development of the photographic method of recording the tracks.

When brought to rest in a photographic emulsion, the positive π-particles decay with the emission of a μ-meson of mass 212 m_e (see Fig. 5). This particle commonly emerges with a constant velocity, so that its range varies only within the narrow limits due to straggling. It follows that in the transmutation of the π-meson, the μ-meson is accompanied by the emission of a single neutral particle. It has now been shown that this neutral particle is of small rest-mass and that it is not a photon. It is therefore reasonable to assume, tentatively, that it is a neutrino, the particle emitted in the process of nuclear β-decay.

When a negative π-meson is arrested in a solid material, it is captured by an atom, interacts with a nucleus and disintegrates it (Perkins[7]; Occhialini and Powell[8]). It follows that the particle has a strong interaction with nucleons, and in this respect its properties are similar to those predicted for the «heavy quanta» of Yukawa.

When the π-mesons are created in nuclear collisions occurring in the atmosphere, they commonly transform, whilst in flight, into μ-mesons and neutrinos. It is these μ-particles which form the «hard» or «penetrating» component of the cosmic radiation and they are responsible for most of the residual ionization in air at sea level. The μ-mesons are penetrating because, unlike the π-mesons, they are able to traverse the nuclei with which they collide without interacting with them, and some of them reach great depths underground.

The production of mesons by protons and α-particles of great energy in nuclear encounters appears to be a result of interactions between nucleons. Accordingly, the heavy nuclei of the primary radiation – if of sufficient energy – also produce similar effects when they collide with other nuclei (see Fig. 6). Because of the large numbers of nucleons involved in such an encounter, the number of mesons produced may be very great.

In addition to producing the charged π-mesons, the primary protons in making nuclear collisions also produce neutral π-particles (Bjorklund *et al.*[9]) The neutral π-mesons are very short-lived and each transforms spontaneously into two quanta of radiation (see Fig. 7). Such a quantum, when it

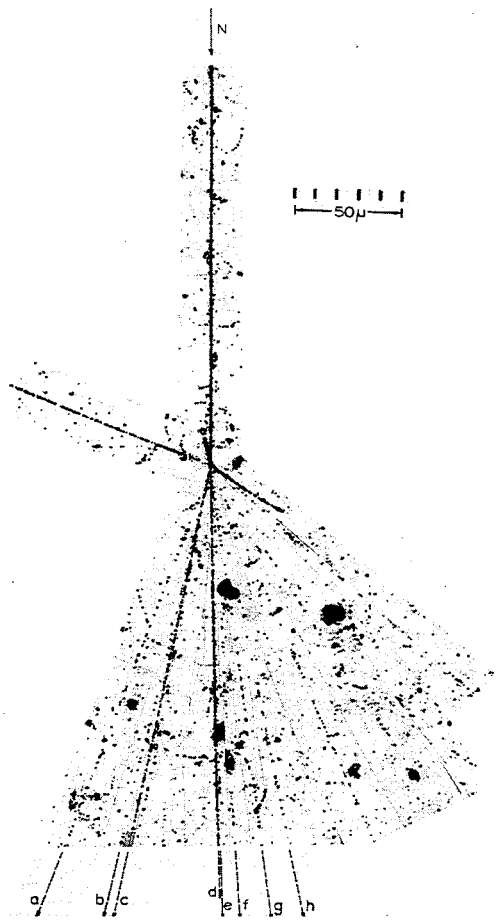


Fig. 6. A nitrogen nucleus of the primary cosmic radiation collides with a nucleus, and splits up into a lithium nucleus, a deuteron, and a number of protons. Some π-mesons are created in the collision. The event is very exceptional in that, by chance, the tracks in the emulsion of many of the particles are long so that they can be identified and their energy determined. *Top:* N, $\bar{\alpha} = 0.0061 \pm 16\%$, $\sim$ 28,000 MeV. *Bottom:* (*a*) π-particle, $\bar{\alpha} = 0.109 \pm 12\%$, 220 MeV; (*b*) π-particle, $\bar{\alpha} = 0.035 \pm 25\%$, 800 MeV; (*c*) ${}_1^3\mathrm{H}$, $\bar{\alpha} = 0.206 \pm 20\%$, 81 MeV; (*d*) Li, $\bar{\alpha} = 0.0075 \pm 30\%$, $\sim$ 10,000 MeV; (*e*) ${}_1^3\mathrm{H}$, $\bar{\alpha} = 0.012 \pm 30\%$, 2000 MeV; (*f*) ${}_1^2\mathrm{D}$, $\bar{\alpha} = 0.0053 \pm 16\%$, 4800 MeV; (*g*) π-particle, $\bar{\alpha} = 0.029 \pm 20\%$, 1000 MeV; (*h*) π-particle, $\bar{\alpha} = 0.025 \pm 30\%$, 1200 MeV.

happens to pass near an atomic nucleus, can in turn transform into a pair of electrons, one positive and one negative; and the electrons can generate new photons in further collisions. A succession of such processes results in the production of the well-known cascades of electrons and photons which form the « soft » or easily absorbed component of the cosmic radiation (Carlson *et al.*[10]).

Although much longer-lived than the π-mesons, some of the μ-mesons also decay in flight to produce electrons and neutrinos. The electrons contribute to the soft component, whilst the neutrinos join the similar particles arising from the decay of the π-mesons to produce a flux of neutral radiation which has a very weak interaction with matter, and of which the fate is at present unknown.

In addition to the π- and μ-mesons, recent experiments at Manchester (Rochester and Butler[11]) and Bristol (Brown *et al.*[12]), and in other laboratories, have shown that more massive types of mesons exist (Fig. 8). Although they occur much less frequently than the π-mesons, the elucidation of their properties appears to be of great importance for the development of nuclear physics.

We are only at the beginning of our penetration into what appears to be a rich field of discovery. Already, however, it seems certain that our present theoretical approach has been limited by lack of essential information; and that the world of the mesons is far more complex than has hitherto been visualized in the most brilliant theoretical speculations. The fast protons and α-particles generated by the cyclotrons are not sufficiently energetic to produce these more massive mesons, but this may become possible when the proton synchrotrons now under construction come into operation.

Only about twenty-five years have passed since it was generally recognized that part of the residual conductivity of a gas at sea level is due to the arrival from out of space of a radiation of great penetrating power. In the 1928 edition of *Conduction of Electricity through Gases*, J. J. Thomson and G. P. Thomson, commenting on this conclusion, remark that « It would be one of the romances of science if these obscure and prosaic minute leakages of electricity from well-insulated bodies should be the means by which the most fundamental problems in the evolution of the cosmos came to be investigated. »

In the years which have passed, the study of what might, in the early days, have been regarded as a trivial phenomenon has, in fact, led us to the discovery of many new forms on matter and many new processes of funda-

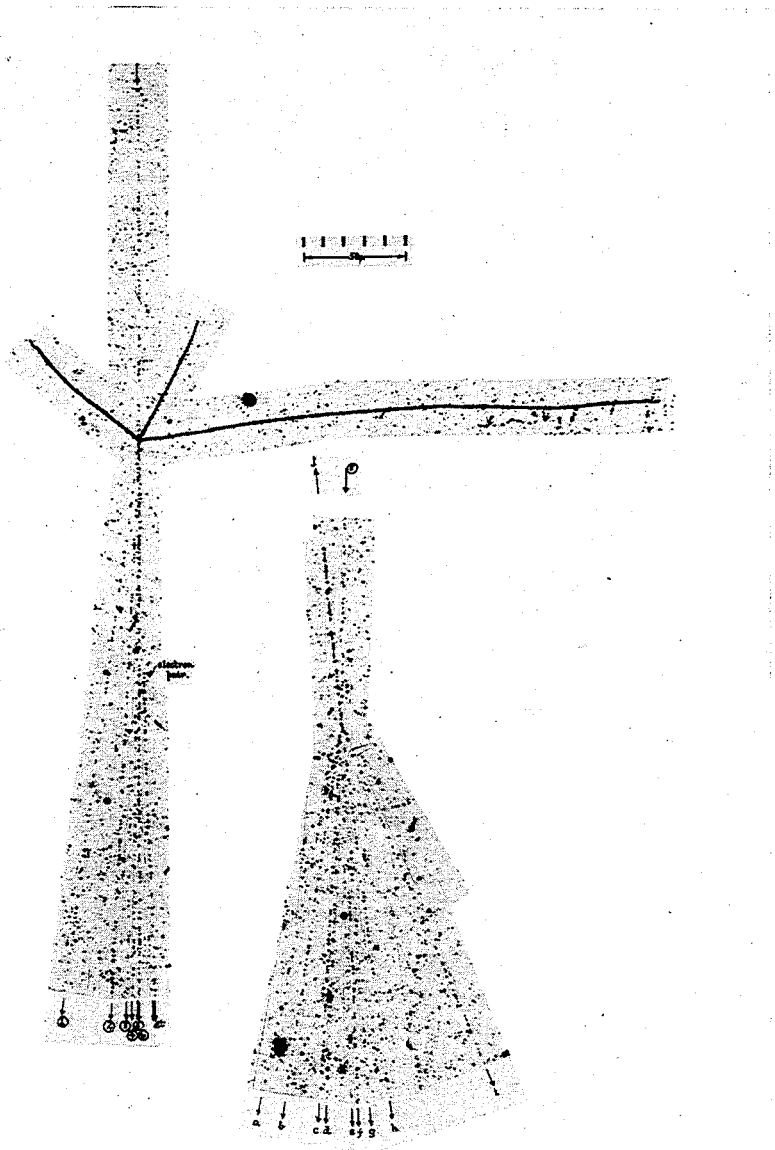


Fig. 7. A fast proton produces a disintegration (*see the left-hand picture*). One of the emerging «shower» particles (No. 5) after passing some millimeters in the emulsion, makes a second disintegration (*see right-hand «star»*).

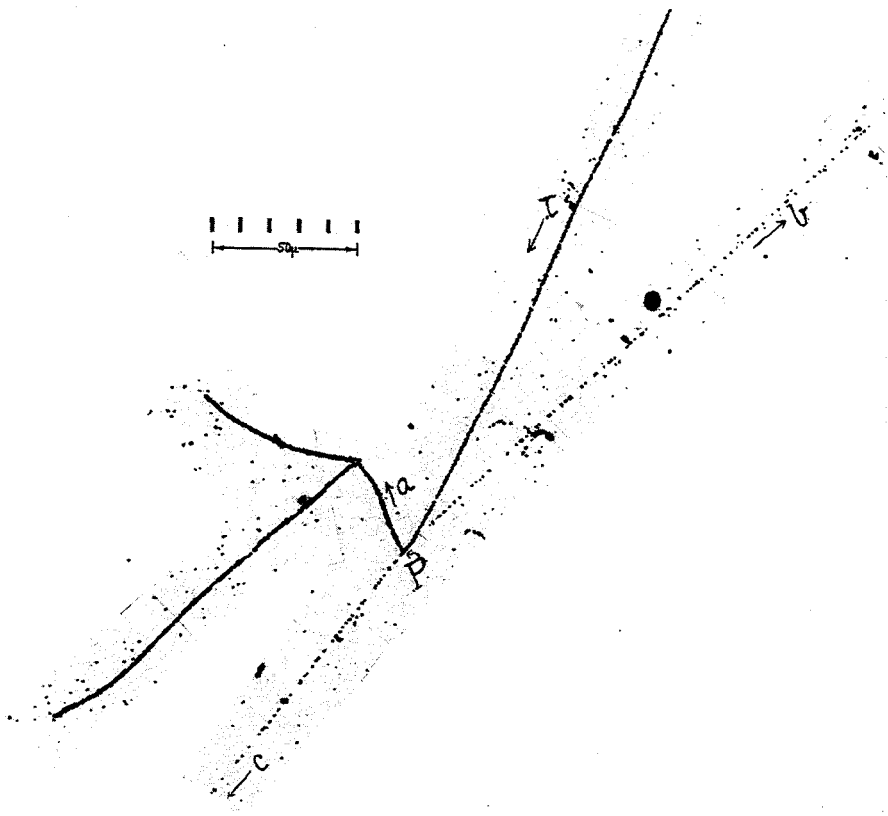


Fig. 8. The particle τ coming to rest in the emulsion at the point P disintegrates into three particles, *a*, *b*, and *c*, of which the initial directions of motion are co-planar. Particle *a* is almost certainly a negative π-meson. Tracks *b* and *c* are long enough to allow the velocity of the corresponding particles to be determined. Particle τ can be proved to be lighter than a proton, but heavier than a π-meson. With the assumption that *b* and *c* are also π-mesons, the momenta of *a*, *b*, and *c* are in balance, without any neutral particle. The τ-particle would then have a mass of about 970 m_e.

mental physical importance. It has contributed to the development of a picture of the material universe as a system in a state of perpetual change and flux; a picture which stands in great contrast with that of our predecessors with their fixed and eternal atoms. At the present time a number of widely divergent hypotheses, none of which is generally accepted, have been advanced to account for the origin of the cosmic radiation. It will indeed be of great interest if the contemporary studies of the primary radiation lead us – as the Thomsons suggested, and as present tendencies seem to indicate – to the study of some of the most fundamental problems in the evolution of the cosmos.

1. P. Freier, E. J. Lofgren, E. P. Ney, F. Oppenheimer, H. L. Bradt, and B. Peters, *Phys. Rev.*, 74 (1948) 213.
2. H. L. Bradt and B. Peters, *Phys. Rev.*, 74 (1948) 1828.
3. C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell, *Nature*, 159 (1947) 694.
4. O. Piccioni, *Phys. Rev.*, 77 (1950) 1, 6.
5. P. H. Fowler, *Phil. Mag.*, 41 (1950) 169.
6. J. R. Richardson, *Phys. Rev.*, 74 (1948) 1720.
7. D. H. Perkins, *Nature*, 159 (1947) 126.
8. G. P. S. Occhialini and C. F. Powell, *Nature*, 159 (1947) 186.
9. R. Bjorklund, W. E. Crandall, B. J. Moyer, and H. F. York, *Phys. Rev.*, 77 (1950) 213.
10. A. G. Carlson, J. E. Hooper, and D. T. King, *Phil. Mag.*, 41 (1950) 701.
11. G. D. Rochester and C. C. Butler, *Nature*, 160 (1947) 855.
12. R. Brown, U. Camerini, P. H. Fowler, H. Muirhead, C. F. Powell, and D. M. Ritson, *Nature*, 163 (1949) 47, 82.

Biography

Cecil Frank Powell was born on December 5th, 1903, at Tonbridge, Kent, where his father, Frank Powell, was one of a family of gunsmiths who had long practised the trade in the town. His grandfather, George Bisacre, had established a private school in the nearby town of Southborough and his family ties and influences therefore tended to encourage a regard for the value both of learning and the practical arts.

He attended a local elementary school and won a scholarship, at the age of eleven, to Judd School, Tonbridge. From there he won open scholarships to Sidney Sussex College, Cambridge, where he graduated with First Class Honours in the Natural Science Tripos (1924–1925).

As a postgraduate student, Powell worked in the Cavendish Laboratory under C. T. R. Wilson and Lord Rutherford until 1927 when he gained his Ph.D. and moved to the University of Bristol as Research Assistant to A. M. Tyndall in the H. H. Wills Physical Laboratory. He was eventually appointed lecturer, then reader and, in 1936, he visited the West Indies as seismologist of an expedition investigating volcanic activity. He returned to Bristol in the following year and in 1948 he was established as Melville Wills Professor of Physics.

Powell was Director of a European expedition for making high-altitude balloon flights in Sardinia (1952) and in the Po Valley (1954, 1955, and 1957).

His first researches at the Cavendish Laboratory concerned condensation phenomena and it led indirectly to an explanation of the anomalously high rate of discharge of steam through nozzles. He showed this to be due to the existence of supersaturation in the rapidly expanding steam and his results were found to have a bearing on the design and performance of the steam turbine.

At Bristol he devoted years of patient work to the development of accurate techniques for measuring the mobility of positive ions and to establishing the nature of the ions in most of the common gases. After his sojourn in the Caribbean, he returned to work on the construction of a Cockcroft generator for accelerating fast protons and deuterons – employing them in con-

junction with a Wilson chamber, to study neutron-proton scattering. In 1938, he undertook experiments in cosmic radiation and employed methods of directly recording the tracks of the particles in photographic emulsions and, when the Cockcroft machine came into operation, he employed similar methods for determining the energy of neutrons, that is, by observing the tracks of the recoiling protons. The length of the track of a charged particle in the emulsion was found to give an accurate measure of its range and the great advantages of this method for experiments in nuclear physics were soon clearly established.

This development led him to a study of the scattering and disintegration processes produced by a beam of high-energy deuterons and he later returned, with the development of photographic emulsions of increased sensitivity, to experiments on cosmic radiation: in 1947 heavy mesons were discovered and many of their more important properties established.

Powell has contributed numerous papers to learned societies on the discharge of electricity in gases, and on the development of photographic methods in nuclear physics. He is a co-author of *Nuclear Physics in Photographs* (1947) and *The Study of Elementary Particles by the Photographic Method* (1959).

Prof. Powell was elected Fellow of the Royal Society in 1949: he was awarded the Hughes Medal in the same year and the Royal Medal in 1961. He has received honorary Doctor of Science degrees from the Universities of Dublin, Bordeaux and Warsaw, and he is a Foreign Member of the Academy of Sciences of the U.S.S.R. He was Vernon Boys Prizeman and Honorary Fellow of the Physical Society (1947), and he served on the Scientific Policy Committee of the European Organization for Nuclear Research (Geneva, 1961).

Powell married Isobel Therese Artner, who has assisted him in his researches, in 1932; they have two daughters. His chief recreations are squash racquets and tennis.

Physics 1951

Sir JOHN DOUGLAS COCKCROFT

ERNEST THOMAS SINTON WALTON

«for their pioneer work on the transmutation of atomic nuclei by artificially accelerated atomic particles»

Physics 1951

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

Your Majesties, Royal Highnesses, Ladies and Gentlemen.

By giving the Nobel Prize in Physics of this year to Sir John Cockcroft, Director of the Atomic Energy Research Establishment at Harwell, and Professor Ernest Walton of Dublin University, the Swedish Academy of Sciences has rewarded a discovery which stands out as a milestone in nuclear research.

At the beginning of this century, the study of the naturally radioactive substances had shown that their property of emitting radiation is connected with spontaneous transmutations of their atoms. It appeared, however, to be beyond human power to influence the course of these processes.

The radiation emitted by a radium source contains swiftly moving and positively charged helium atoms. By investigating the way in which these particles are deflected by other atoms, the great nuclear scientist Rutherford found in 1911 that an atom has a positive nucleus which is very small compared to the whole atom but contains most of its mass. Besides the nucleus, the atom contains negative electrons, moving around the nucleus.

Continuing these investigations, Rutherford was able in 1919 to produce transmutations of atomic nuclei by bombarding nitrogen with helium nuclei from a radium source. Some of the helium nuclei had enough energy to overcome the repelling electric field and to penetrate into the nitrogen nucleus, in those rare cases when they struck such a nucleus. The nitrogen nucleus thereupon turned into an oxygen nucleus, while a hydrogen nucleus was emitted.

Thus it became possible by external means to transform nitrogen into oxygen, i.e., to transmute one element into another.

However, only a very few nuclear transmutations could be produced by these natural projectiles, the helium nuclei from radioactive substances. In order to produce nuclear transmutations on a larger scale, and thus obtain further insight into the structure of atomic nuclei, a more powerful stream of projectiles was needed.

Accordingly, the end of the 1920's saw investigations of the possibility

of accelerating charged particles to high energies, with the ultimate aim of using these particles to produce nuclear transmutations. This year's Nobel Laureates in Physics were the first to succeed in this task, by their joint work at the Cavendish Laboratory in Cambridge, of which Rutherford was at that time the director. In planning this work, they realized the importance of certain contemporary theoretical studies by Gurney and Condon, and by Gamow. This work had shown that, because of the wave properties of matter, there is a certain probability for a positively charged particle to penetrate into a nucleus even if, according to ordinary mechanical concepts, the velocity of the particle does not suffice to overcome the electric repulsion from the nucleus. Cockcroft had emphasized that the conditions are particularly favourable if hydrogen nuclei are used as projectiles, and that an accelerating voltage of only a few hundred thousand volts should suffice to give observable transmutations of light elements.

The work of Cockcroft and Walton was a bold thrust forward into a new domain of research. Great difficulties had to be overcome before they were able to achieve their first successful experiments at the beginning of 1932. By then, they had constructed an apparatus which, by multiplication and rectification of the voltage from a transformer, could produce a nearly constant voltage of about six hundred thousand volts. They had also constructed a discharge tube in which hydrogen nuclei were accelerated. Causing these particles to strike a lithium layer, Cockcroft and Walton observed that helium nuclei were emitted from the lithium. Their interpretation of this phenomenon was that a lithium nucleus into which a hydrogen nucleus has penetrated breaks up into two helium nuclei, which are emitted with high energy, in nearly opposite directions. This interpretation was later fully confirmed.

Thus, for the first time, a nuclear transmutation was produced by means entirely under human control.

In order to get a detectable transmutation of lithium, a voltage of little more than one hundred thousand volts was required. The number of transmutations rose quickly as the voltage was increased. The corroboration obtained in this way for the theory which Gamow and others had propounded, and which was referred to above, was of great importance.

The analysis made by Cockcroft and Walton of the energy relations in a transmutation is of particular interest, because a verification was provided by this analysis for Einstein's law concerning the equivalence of mass and energy. Energy is liberated in the transmutation of lithium, because the total

kinetic energy of the helium nuclei produced is greater than that of the original nuclei. According to Einstein's law, this gain in energy must be paid for by a corresponding loss in the mass of the atomic nuclei. This assertion was satisfactorily confirmed by Cockcroft and Walton, experimental errors being taken into consideration. Somewhat later, more exact investigations based on the same principles gave a complete verification of Einstein's law. Thus a powerful method was obtained for comparing masses of atomic nuclei.

In subsequent work, Cockcroft and Walton investigated the transmutations of many other atomic nuclei. Their techniques and results remain a model for nuclear research. As projectiles, they also used the nuclei of heavy hydrogen, which had then just been discovered. As end products, several atomic nuclei were obtained which had not been known previously. Following the discovery of artificially radioactive elements, by Frédéric and Irène Joliot-Curie, they found that such elements can also be produced by irradiation with hydrogen nuclei.

The investigations of Cockcroft and Walton disclosed a new and fertile domain of research, consisting of the study of nuclear transmutations of various types.

Their discoveries initiated a period of rapid development in nuclear physics. Besides the apparatus of Cockcroft and Walton, the cyclotron constructed by Lawrence, and various other particle accelerators played important roles. By its stimulation of new theoretical and experimental advances, the work of Cockcroft and Walton displayed its fundamental importance. Indeed, this work may be said to have introduced a totally new epoch in nuclear research.

Sir John Cockcroft, Professor Ernest Walton. The great nuclear scientist Rutherford, with whose work your discovery is closely connected, sometimes used to say: « it is the first step that counts ». This saying may be applied in the truest sense to your discovery of the transmutations of atomic nuclei by artificially accelerated particles. Indeed, this work of yours opened up a new and fruitful field of research which was eagerly seized upon by scientific workers the world over. It has profoundly influenced the whole subsequent course of nuclear physics. It has been of decisive importance for the achievement of new insight into the properties of atomic nuclei, which could not even have been dreamt of before. Your work thus stands out as a landmark in the history of science.

On behalf of the Royal Swedish Academy of Sciences may I extend to you our warmest congratulations. I now ask you to receive your Nobel Prize from the hands of His Majesty the King.

JOHN D. COCKCROFT

Experiments on the interaction of high-speed nucleons with atomic nuclei

Nobel Lecture, December 11, 1951

The experimental researches in nuclear physics with which I have been associated have been concerned with the transmutation of atomic nuclei, the study of their level structure and of the forces which hold nuclei together. I began this work in the Cavendish Laboratory under the direction of Lord Rutherford in 1928. At this time experimental work on the energies of α-particles ejected from the radioactive elements had shown that these particles could have a substantially lower energy than the calculated height of the potential barrier around the nucleus. For a time this was somewhat of a puzzle, but in 1928 Gamow, who was then working in Copenhagen, and also Gurney and Condon, showed that this could be readily explained by attributing wave properties to the escaping α-particle so that the particle could escape from the nucleus without having a high enough energy to surmount the potential barrier. Gamow visited the Cavendish Laboratory in 1928 and I discussed with him the converse problem – the energy which would be required for a proton accelerated by high voltages to penetrate the nuclei of the light elements. As a result of these talks I prepared a memorandum which I sent to Rutherford showing that there was a quite high probability for the boron nucleus to be penetrated by a proton of only 300 kilovolts energy whilst conditions for lithium were even more favourable. Rutherford then agreed to my beginning work on this problem and I was soon joined by Dr. Walton who had previously been working on the development of an early linear accelerator and also on an equally early betatron.

Dr. Walton will describe the work on the development of our equipment. With our first apparatus we produced beams of protons from a canal ray tube and accelerated them by voltages of up to 280 kilovolts, and in 1930 bombarded a lithium target and other targets and looked for γ-rays with a gold-leaf electroscope. We found then only soft continuous radiation; we did not find gamma rays – as we now know, our proton energy was too far below the first resonance which gives rise to gamma ray emission. We were

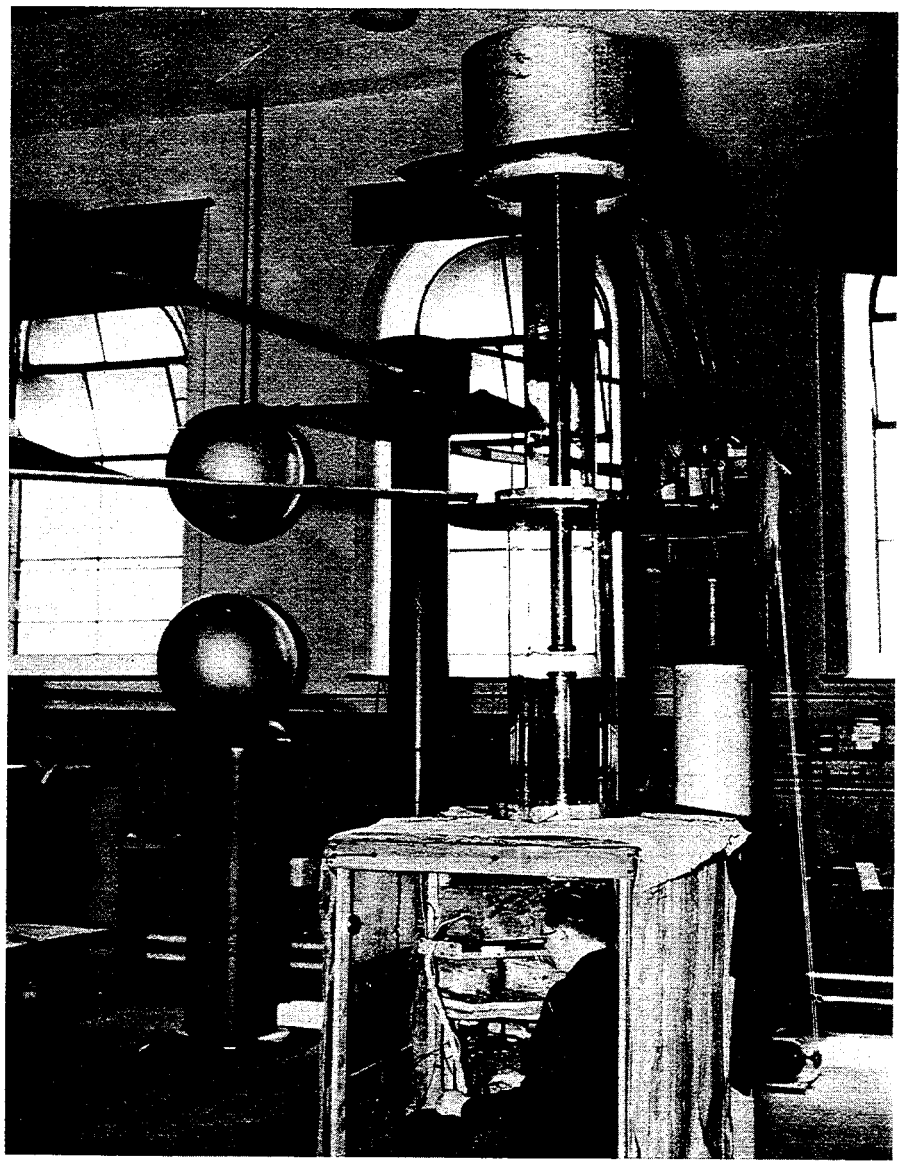


Fig. 1 a.

then interrupted by the necessity to move our laboratory and in doing so we decided to increase the energy of our protons.

Fig. 1 a and 1 b show the new apparatus which was completed in early 1932. We were soon able to bring a narrow beam of 500 kilovolts protons

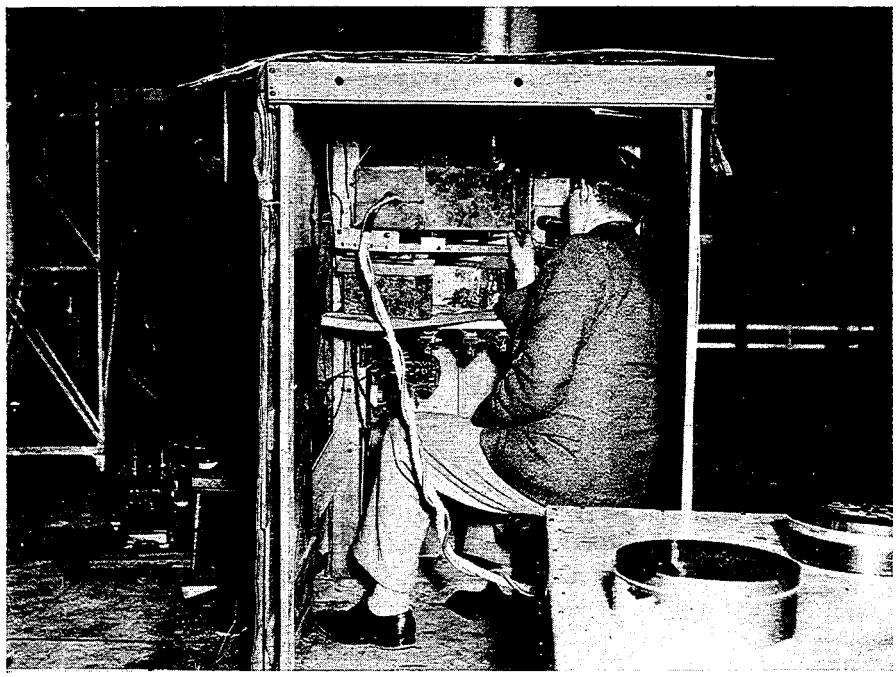


Fig. 1 b.

out through a thin mica window in the base of the experimental tube, and to measure their range as a function of energy.

Soon after this, we resumed our experiments on lithium, but this time, instead of looking for gamma rays, we set out to look for α-particles from the disintegration of lithium. A mica window was provided to allow the α-particles to escape, and opposite the mica window we placed the well-tried tool of Rutherford – the zinc sulphide screen (Fig. 2). Almost at once, at an energy of 125 kilovolts, Dr. Walton saw the bright scintillations characteristic of α-particles, and a first primitive absorption experiment showed that they had a range of about 8.4 cm. We then confirmed by a primitive coincidence experiment, carried out with two zinc sulphide screens and two observers tapping keys (Fig. 3), that the α-particles were emitted in pairs. Our resolving time was a second or so – somewhat longer than the resolving time of modern coincidence circuits which operate in units of millimicroseconds. More refined experiments showed that the energy of the α-particles was 8.6 million volts (Fig. 4). It was obvious then that lithium was being disintegrated into two α-particles with a total energy release of 17.2 million volts.

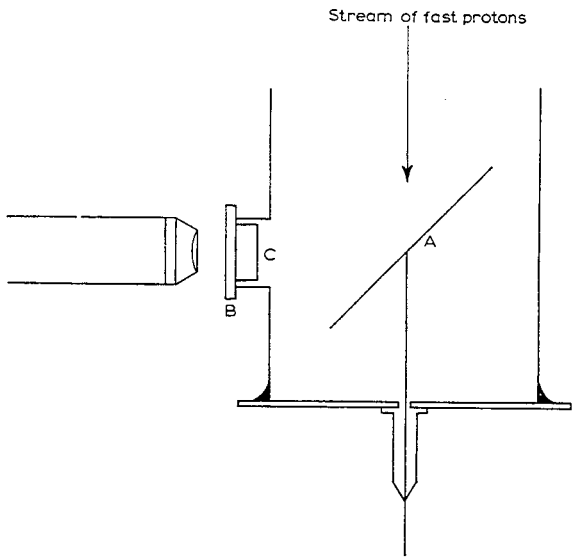


Fig. 2.

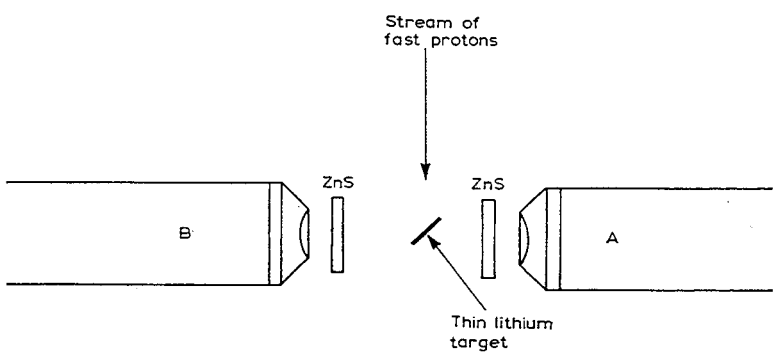


Fig. 3.

This energy could be provided by a diminution of mass of 0.0184 mass units.

The mass balance of the reaction at that time was

^{7}Li	7.0104 (Costa)
^{1}H	1.0072
	8.0176
2 ^{4}He	8.0022
Mass decrease	0.0154

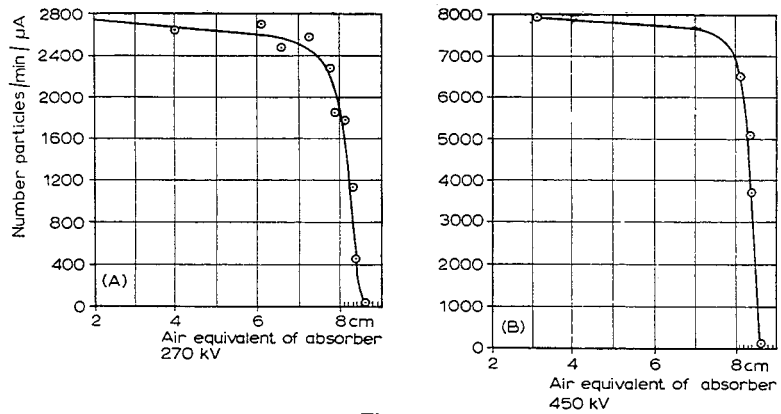


Fig. 4.

A little later Bainbridge redetermined the mass of ^{7}Li to be 7.0130. This changed the mass decrease to 0.0180 mass units, in very good agreement with the observed figure.

We also studied the variation of the number of disintegrations with energy and obtained the results shown in Fig. 5. The increase in disintegration with higher proton energy results from the increasing probability of penetration of the nuclei potential barriers predicted by the Gamow, Gurney and Condon theory.

We studied also the disintegration of boron and found a different type of absorption curve (Fig. 6) for the α-particles which were emitted with a continuous distribution in energy with a maximum range of 4.4 cm. The continuous distribution in energy was explained by boron breaking up into α-particles. We found also that α-particles were emitted from most elements but we found later that these were due largely to boron impurities in our targets, boron having a very high probability of disintegration.

These experiments were powerfully supported by Dee who was diverted by Rutherford from his cloud chamber work carried out with C. T. R. Wilson to join in this promising new field. Dee, with Walton, built himself another accelerating tube and arranged for protons to pass down a tube into a Wilson chamber. Thin mica windows were arranged for the α-particles to enter the cloud chamber. Figs. 7 and 8 show the transmutation of lithium into two α-particles and of boron into three α-particles.

Our next group of experiments were carried out with ions of heavy hydrogen which became available to the laboratory in 1933 through the kindness of Professor G. N. Lewis who sent a few cubic millimetres of heavy

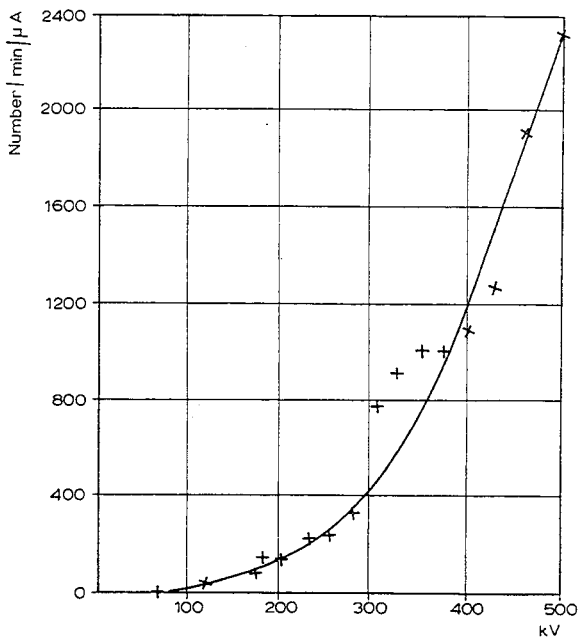


Fig. 5.

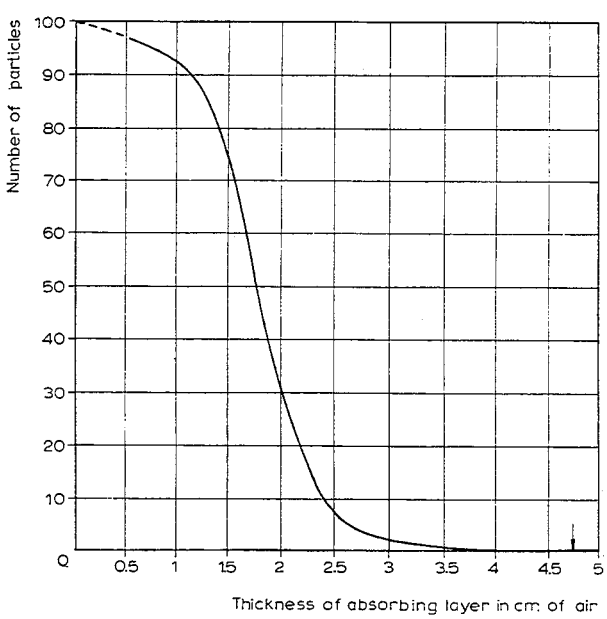


Fig. 6. Absorption curve of boron.

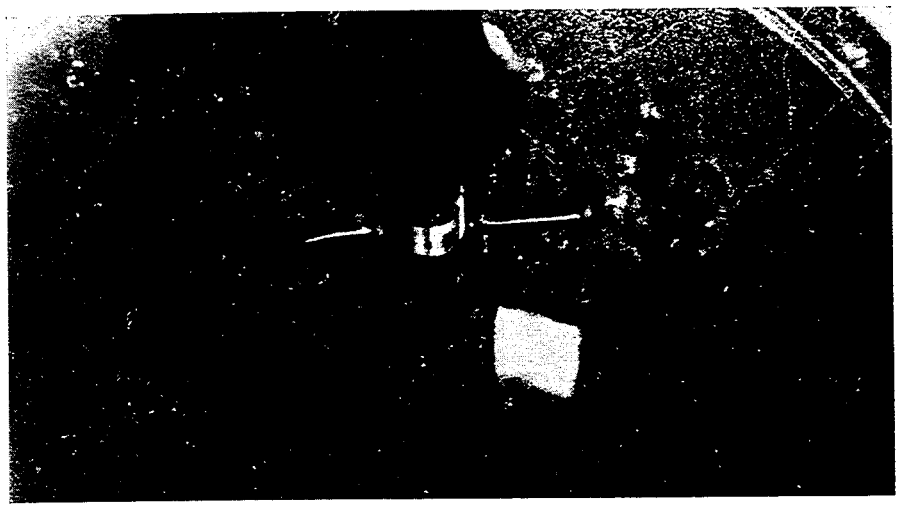


Fig. 7.

water over to Rutherford. Lawrence, Livingston, and Lewis, had carried out some pioneer experiments on disintegration produced by deuterons and had found that a group of protons of 18 cm range was emitted from most targets. By this time Oliphant had built a low-voltage accelerator to enable him to work with Rutherford in this exciting new field. Together they discovered the disintegration of deuterium by deuterons – or diplons as Rutherford wanted to call them. Figure 9 is a beautiful photograph by Dee showing the disintegration into a triton and a proton. The protons explained the observations of Lawrence, Livingston, and Lewis. In September, 1933, I brought back two cans of dilute heavy water from Berkeley and after this was concentrated by Harteck we obtained a little for our own use. We then studied the (d, p) type of reaction – the transmutation of ^{6}Li into ^{7}Li with the emission of a proton; the transmutation of ^{10}B into ^{11}B, ^{12}C into ^{13}C, again with the emission of one or more proton groups. The multiple proton groups showed the existence of excited levels of the residual nucleus.

During the course of these experiments we had looked for the emission of delayed α-particles from nuclei but had never found them. In early 1934 we heard of the production of artificial radioactivity by Curie and Joliot who observed a delayed emission of positrons from targets of boron, magnesium, and aluminium bombarded by α-particles. Walton and I were able to borrow a Geiger-counter equipment from Dr. Bainbridge – there was only one portable equipment in the Cavendish Laboratory at that time – and at

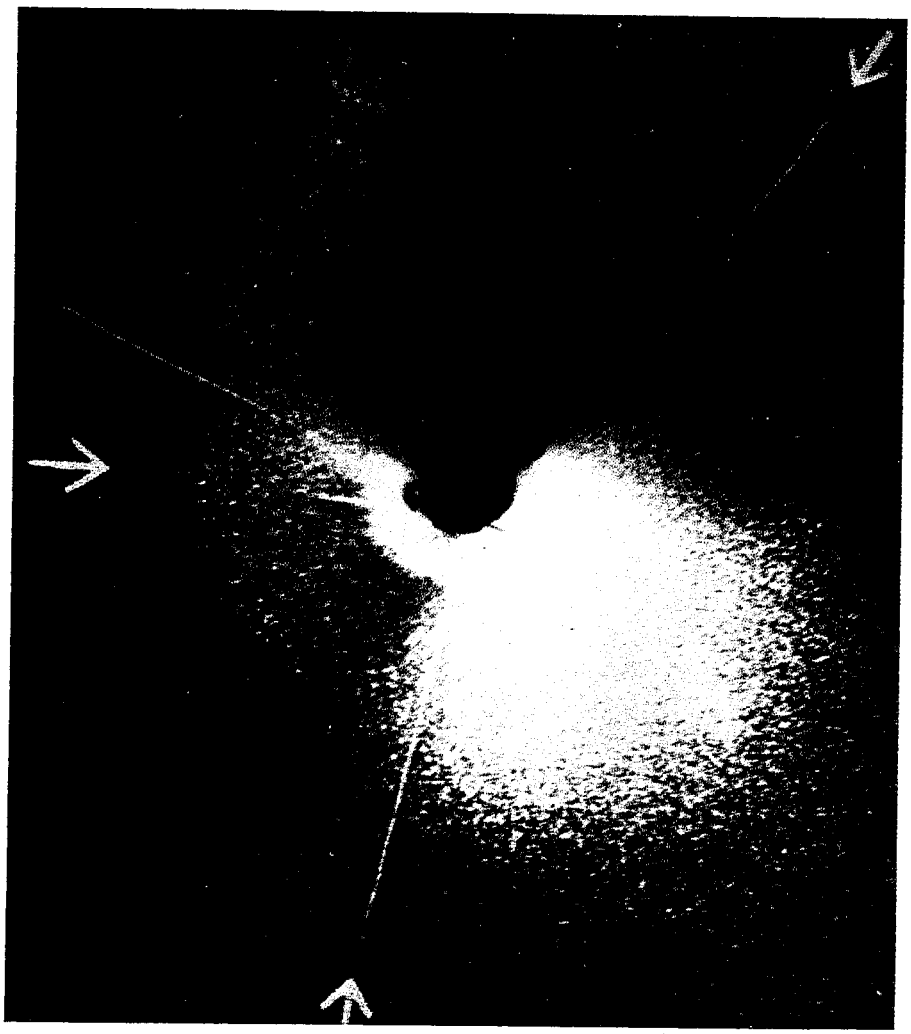


Fig. 8.

once found that when graphite was bombarded by a beam of 400–500 kilovolt protons, delayed positrons were emitted, and that the activity had a half-life of about 11 minutes. We found also that similar effects were produced by the deuteron bombardment of carbon, an effect which had also been found by Lauritsen and Crane. These two reactions were seen to be due to the formation of nitrogen 13, the first by proton capture and the second by a (d,p) reaction.

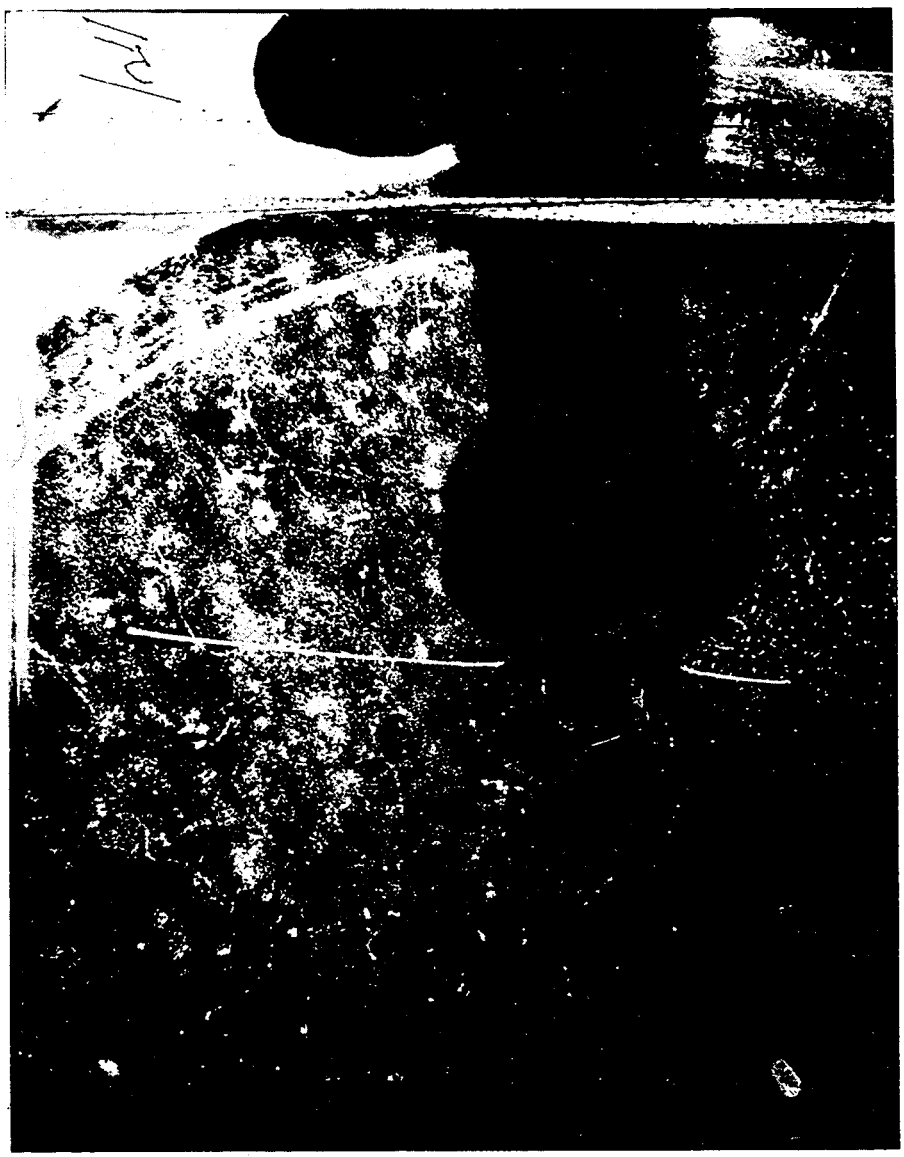


Fig. 9.

The next group of experiments were carried out in association with Dr. W. B. Lewis. Our apparatus was greatly improved in accuracy and we were able to study the disintegration of boron, carbon, nitrogen, and oxygen by deuterons with some precision.

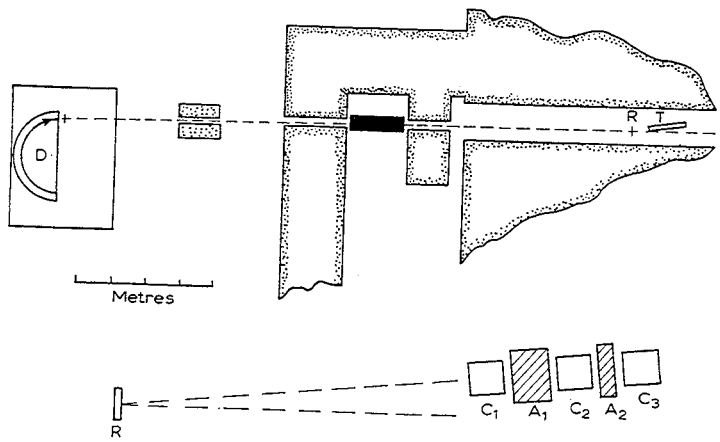


Fig. 10.

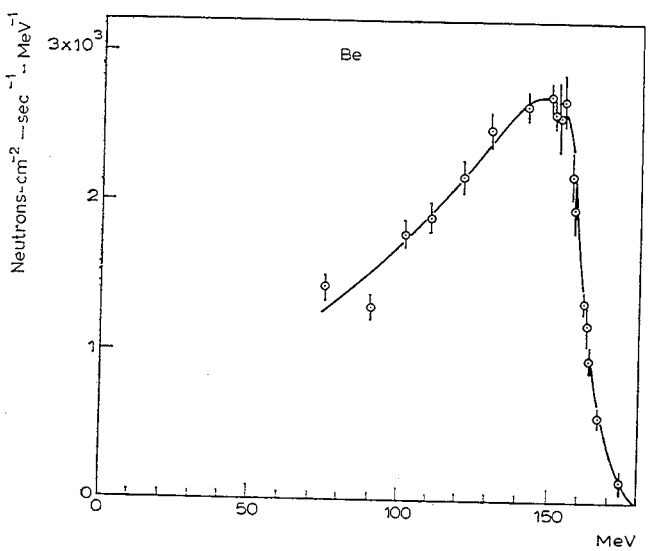


Fig. 11.

We studied for example the interesting (d,α) reactions in carbon and nitrogen

$$^{12}_{6}\mathrm{C} + ^{2}_{1}\mathrm{D} \rightarrow ^{10}_{5}\mathrm{B} + ^{4}_{2}\mathrm{He}$$

$$^{14}_{7}\mathrm{N} + ^{2}_{1}\mathrm{D} \rightarrow ^{12}_{6}\mathrm{C} + ^{4}_{2}\mathrm{He} \ [\text{or} \rightarrow 4\ ^{4}_{2}\mathrm{He}]$$

These new reactions enabled a close check to be made on the new scale of nuclear masses which had recently been proposed by Bethe and Oliphant.

For some time after this I was diverted by Rutherford from nuclear physics to take charge of the Royal Society Mond Laboratory when Kapitza was retained in Russia. I was interested however to some extent in the building of the Cavendish Laboratory cyclotron, which I was at last able to persuade Rutherford to build, on my return from Berkeley in 1937; and in the High Voltage Laboratory, with its voltage multiplier accelerator for 1 million volts. Dee was however the effective leader of this laboratory.

After the War in the autumn of 1946 with many colleagues I started to transform the Royal Air Force Station at Harwell into a Research Establishment. Since then we have built a synchrocyclotron for producing 180 MeV protons, a Van de Graaff generator for experiments with 3 MeV protons

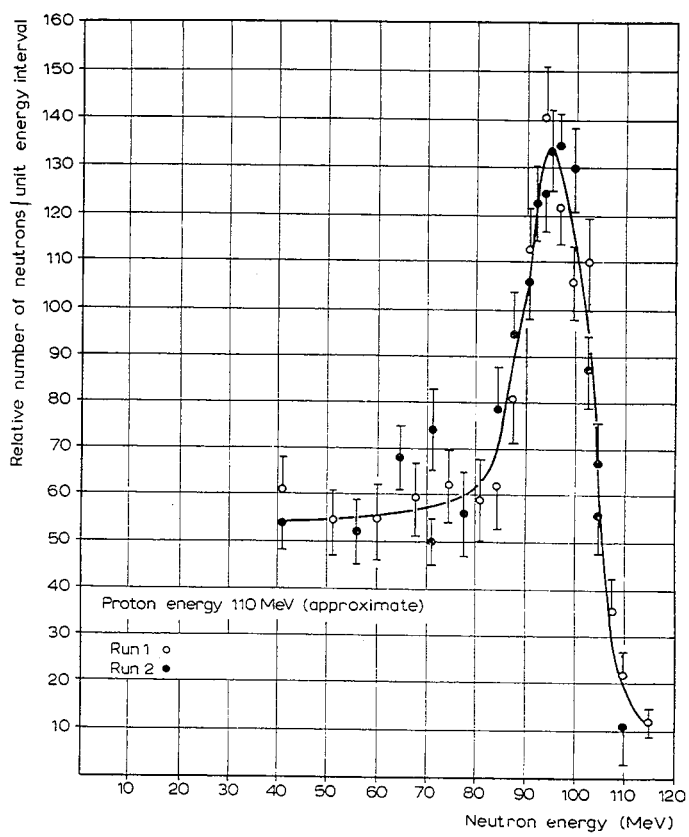


Fig. 12. Neutron energy distribution for beryllium.

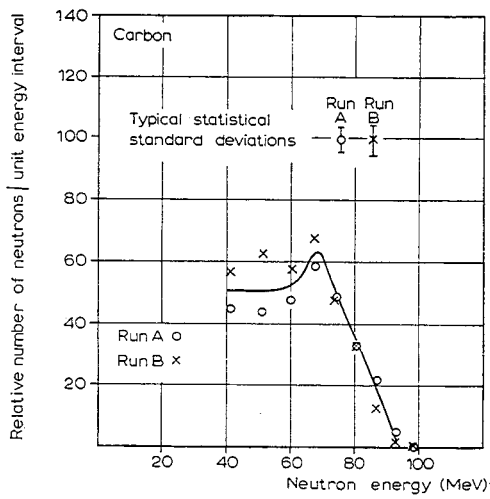


Fig. 13. Energy distribution of neutrons from carbon target bombarded by 100-MeV protons.

and deuterons, and a linear accelerator for electrons as our principle tools of nuclear physics. I will now describe some of the experiments which have been carried out by the cyclotron group working under the leadership of Dr. Pickavance.

Experiments on nuclear forces have been carried out with beams of neutrons and protons. High-energy neutrons are produced by allowing the proton beam to strike a target inside the vacuum tank. Neutrons which are projected forward pass through a thin window in the wall of the vacuum tank and then through collimating tubes in blocks of concrete having a total thickness of $4\frac{1}{2}$ metres. (Fig. 10). The energy of the neutrons can be determined by allowing them to project protons from a polythene disc. The projected protons then pass through a coincidence «telescope» of two proportional counters and one gridded ionization chamber which records slow protons. Fig. 11 shows the energy spectrum of the neutrons from beryllium. A fairly high proportion of the neutrons are projected forward with almost the full energy of the protons. The tail of neutrons is due to multiple scattering inside the nucleus. I will also show the results of a similar experiment (Fig. 12) carried out by the Harvard workers using 110 MeV protons. You will see that the peak is more pronounced. It seems that nuclear matter is particularly transparent to neutrons of about 110 MeV energy, due to a peculiar interference effect. You will perhaps think that the Harvard peak

is more clearly defined because of better geometry. So I will show you their results for carbon (Fig. 13). The peak has now almost vanished. This must be connected with the fact that about 20 MeV of energy are required to extract a neutron from carbon so that the neutrons will be of a lower energy for which the nuclear matter is not so transparent. Fig. 14 shows the Harwell results for aluminium. You will see that there is now a very pronounced tail. This must be due again to the neutrons losing energy making multiple small-angle collisions in the nuclear matter before emerging.

Experiments have also been carried out by Mrs. Skyrme to investigate the spectrum of lower energy neutrons. Fig. 15 shows that there is a strong group of « evaporation neutrons » which have a peak energy of about 1 MeV. It appears that about four neutrons are evaporated from a Tungsten target for each inelastic collision of a high-energy neutron with a target. Smaller numbers of neutrons are evaporated from light nuclei.

Experiments have also been carried out on neutron–proton scattering. The beam of neutrons was used to project protons, and the number projected at different angles was measured. The effective neutron energy could be defined by two limits – the first set by only recording protons through a given thickness of graphite – the second from the maximum energy of the internal proton beam in the cyclotron.

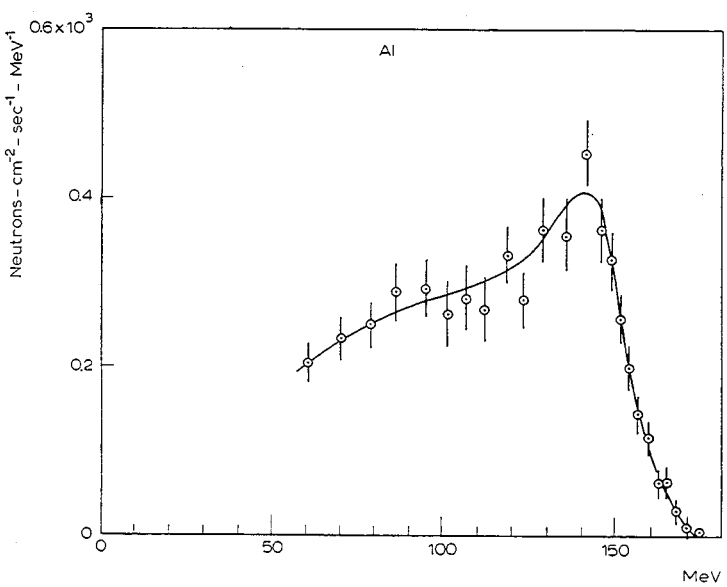


Fig. 14.

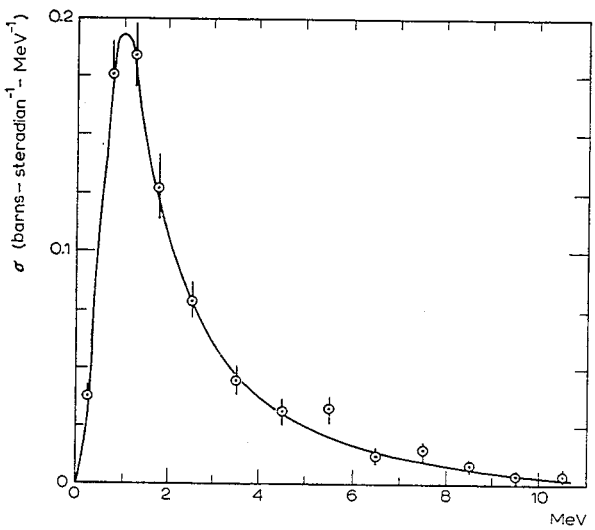


Fig. 15.

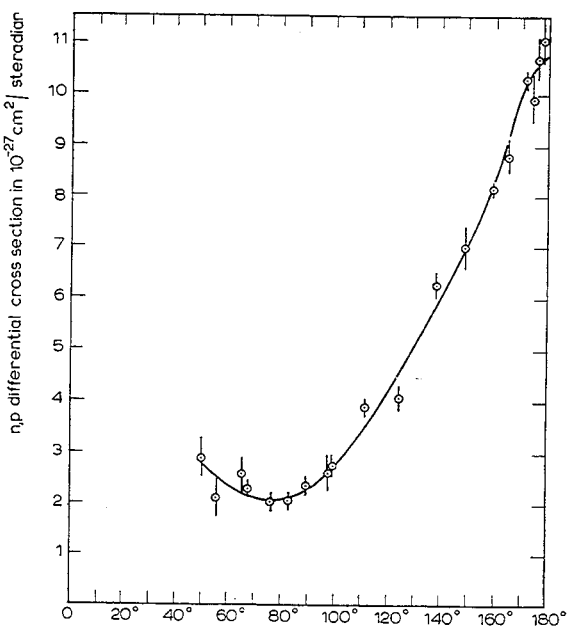


Fig. 16. *Abscissa:* Neutron scattering angle.

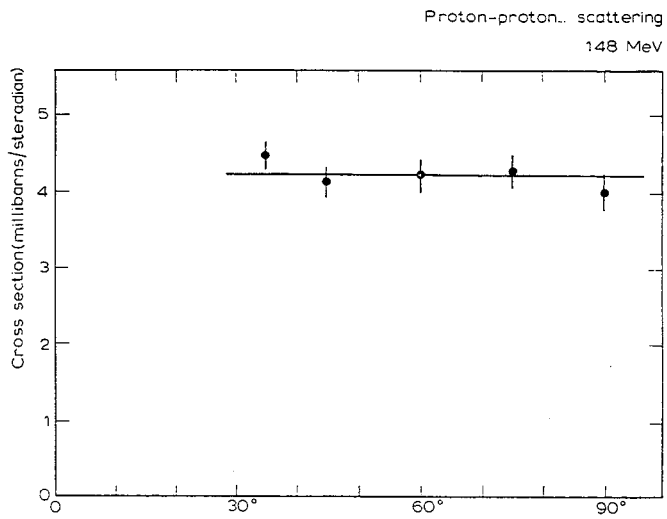


Fig. 17. *Abscissa:* Centre of mass scattering angle.

The results of the (n,p) scattering experiment, plotted in centre of gravity co-ordinates are shown in Fig. 16. You will see that there is a pronounced minimum and that the general character of the scattering is far removed from elastic sphere scattering.

Experiments were then carried out on (p,p) scattering. The protons are scattered out of their final orbit by a uranium foil and then enter a magnetically screened channel so that they can escape from the cyclotron into a long pipe. We obtain about 10^5 protons /cm^2 per second at a distance of 40 feet. The protons, which have an energy of about 145 MeV are then scattered by polythene and the projected protons are received by a telescope of counters. Background effects can also be reduced by recording the scattered proton at 90°. The results are shown in Fig. 17. You will see that the (p,p) cross section is almost independent of angle in centre of gravity co-ordinates down to 30°. The characteristics of (p,p) scattering are therefore quite different from (p,n) scattering. Some differences would be expected owing to the effect of the Pauli exclusion principle on (p,p) scattering. It is therefore not certain that the (p,p) and (p,n) forces are basically different.

The results for the (p,p) differential scattering cross section at 90° can be compared with the values obtained at other energies of other cyclotron groups. The cross section tends to become constant at energies above about 200 MeV.

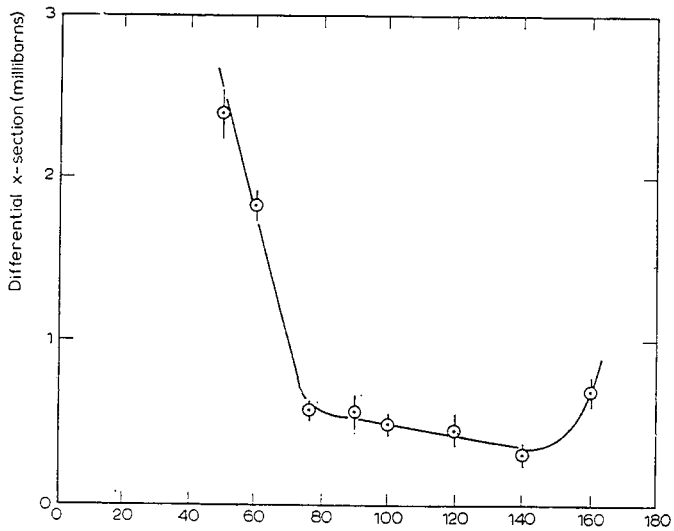


Fig. 18. *Abscissa:* Centre of mass angle.

A further experiment was carried out on the inelastic scattering of protons by deuterons. To achieve this, the protons were scattered from heavy water. The results of this experiment are shown in Fig. 18. It is hoped next to carry out a similar experiment on the scattering of neutrons by deuterons.

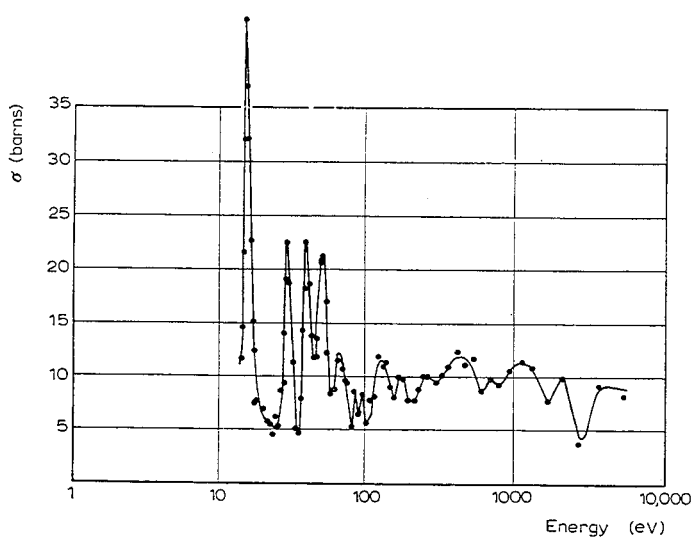


Fig. 20. Total cross section of silver.

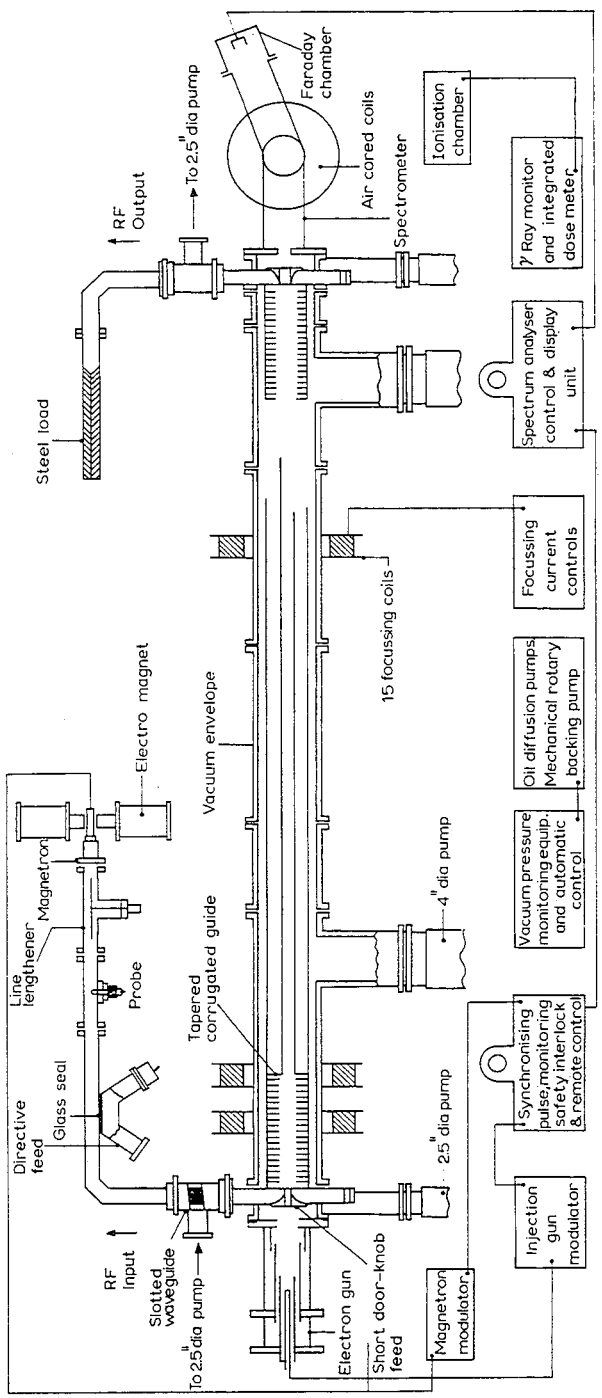


Fig. 19. Schematic diagram of 4.0 MeV linear accelerator.

The linear accelerator. During the post-war years another group at Harwell has developed the electron travelling wave linear accelerator. Fig. 19 shows a diagram of the 3.5 MeV accelerator. This accelerator has been used at Harwell to produce an intense pulsed source of neutrons. The electrons are produced in pulses having a duration of about two microseconds and produce high energy X-rays in a heavy target which in turn splits up heavy water and produces neutrons. About 2.10^{12} neutrons/second are produced in the pulse when the beam current in the accelerator is 120 milliamperes, when the electron energy is 3.2 MeV. The neutrons are partially slowed down in the heavy water. We then use the time of flight method to carry out experiments on neutrons of a given energy. Fig. 20 shows results for the total cross section of silver.

The energy of this linear accelerator is now being extended to 13 MeV and perhaps to 15 MeV. We then hope to increase the neutron intensity by a factor of 100. We may also reduce the pulse width to a fraction of a microsecond. After this we hope to have good resolution up to 10,000 volts.

Work is also proceeding on linear accelerators to produce much higher energy particles.

Biography

John Douglas Cockcroft was born at Todmorden, England, on May 27th, 1897. His family had for several generations been cotton manufacturers.

He was educated at Todmorden Secondary School and studied mathematics at Manchester University under Horace Lamb in 1914–1915. After serving in the First World War in the Royal Field Artillery he returned to Manchester to study electrical engineering at the College of Technology under Miles Walker. After two years apprenticeship with Metropolitan-Vickers Electrical Company he went to St. John's College, Cambridge, and took the Mathematical Tripos in 1924. He then worked under Lord Rutherford in the Cavendish Laboratory.

He first collaborated with P. Kapitza in the production of intense magnetic fields and low temperatures. In 1928 he turned to work on the acceleration of protons by high voltages and was soon joined in this work by E. T. S. Walton. In 1932 they succeeded in transmitting lithium and boron by high energy protons. In 1933 artificial radioactivity was produced by protons and a wide variety of transmutations produced by protons and deuterons was studied. In 1934 he took charge of the Royal Society Mond Laboratory in Cambridge.

In 1929 he was elected to a Fellowship in St. John's College and became successively University demonstrator, lecturer and in 1939 Jacksonian Professor of Natural Philosophy.

In September 1939 he took up a war-time appointment as Assistant Director of Scientific Research in the Ministry of Supply and started to work on the application of radar to coast and air defence problems. He was a member of the Tizard Mission to the United States in the autumn of 1940. After this he was appointed Head of the Air Defence Research and Development Establishment. In 1944 he went to Canada to take charge of the Canadian Atomic Energy project and became Director of the Montreal and Chalk River Laboratories until 1946 when he returned to England as Director of the Atomic Energy Research Establishment, Harwell.

For the period 1954–1959 he was scientific research member of the U.K.

Atomic Energy Authority and has since continued this function on a part-time basis. Election to Master, Churchill College, Cambridge, followed in October 1959. In addition he is Chancellor of the Australian National University, Canberra, and a past President of the Institute of Physics, the Physical Society (1960 to 1962) and the British Association for the Advancement of Science (1961 to 1963).

He has received honorary doctorates from some 19 universities and is a fellow or honorary member of many of the principal scientific societies. In addition, numerous honours and awards have also been bestowed upon him.

He married Eunice Elizabeth Crabtree in 1925 and has four daughters and a son.

ERNEST T. S. WALTON

The artificial production of fast particles

Nobel Lecture, December 11, 1951

The first successful experiments on the transmutation of atoms were carried out by Rutherford[1] in 1919 using swift naturally emitted α-particles. He found that when nitrogen was thus bombarded, fast protons were ejected. In the decade following, many similar experiments were made and these showed that certain other light elements could be disintegrated in like manner. Further progress was made difficult by the limited strengths of the available radioactive sources and because the energies of the α-particles emitted were too low. It became increasingly obvious that attempts should be made to produce by artificial means suitable streams of fast particles. The energy required would be small – a few microamperes of helium ions accelerated by means of a potential of a few million volts corresponds to the α-particle emission of some hundreds of grams of radium. Further, other particles which are not emitted by any radioactive substance should be accelerated and used in experiments. But the practical difficulties appeared to be immense. The potential barriers surrounding nuclei were known to be of the order of millions of volts in height and such potentials were far beyond anything which had been applied successfully to X-ray or cathode ray tubes. Their use seemed impracticable.

The recognition of these difficulties gave an impetus to the search for methods of producing fast particles without the use of correspondingly high voltages. Two main types may be noticed: (*i*) acceleration by a circular electric field in which the particles circulate many times, and (*ii*) acceleration by a series of impulses given to the particles at suitable regular intervals. These methods will now be described briefly.

(*i*) *The betatron.* This was formerly commonly called an induction accelerator because the circular electric field mentioned in (*i*) above is induced by the variation with time of a magnetic field with axial symmetry (Slepian[2], Wideröe[3]). A charged particle, if constrained to move on the circle, would be accelerated continuously as it travelled round it. If the electric field can be maintained for a sufficiently long time, a light particle could go round the circle many times and thus gain considerable energy. The method is un-

suitable for heavy particles because they do not travel sufficiently fast to traverse the orbit many times while the circular electric field exists. If the particles start from rest when the magnetic field is zero, it so happens that the field varies with time in the correct way to maintain them on the circular orbit. A uniform magnetic field is not suitable, since it is always just twice as strong as is required to constrain the particles to move on a circle, and they would spiral in towards the centre. Early attempts to use the method were not successful because very little consideration was given to the question of the stability of this orbit. This is of the utmost importance of the particle has to traverse the orbit many times. In 1929[4] it was shown that a field which decreased inversely with the distance from the axis of symmetry would constrain the particle to move on a particular circular orbit and, in addition, would give radial stability if certain small time varying electric fields were also present. Later, in their pioneer work on the cyclotron Lawrence and Livingston[5] showed that a magnetic field decreasing with the radial distance would produce axial stability owing to the curvature of the lines of force. The problem was discussed fully by Kerst and Serber[6] in 1941. They showed that both radial and axial stability could be ensured by a field which fell off as $1/r^n$, where n lies between 0 and 1. Kerst's experiments[7] were very successful and his apparatus has developed into the modern betatron which will give electrons of several hundred million electron-volts energy. This seems to be near the useful limit of this method.

(*ii*) *The linear accelerator.* In this method a stream of charged particles is sent down the common axis of a line of cylinders, and accelerated successively by electric fields between adjacent cylinders. If a high-frequency alternating potential is applied between the odd-and-even sets of cylinders, by a suitable choice of the lengths of these cylinders it can be arranged that the field is always an accelerating one when the particle is traversing the gap between a cylinder and the next one.

The potentials on the cylinders are changed while the particles are travelling through them and thus they are shielded from any adverse fields. The method is most suitable for heavy particles because then shorter cylinders and potentials at lower frequency may be used. For the acceleration of particles to high energies, the electrical capacity of the system requires large charging currents. In practice this requires considerable high-frequency power. As the path of a particle is long, only small output currents will be obtained unless good focusing is present. The principle of the method was suggested first by Ising[8] in 1925, but some years had to elapse before tech-

nique was sufficiently advanced for results of practical importance to be obtained. In 1928 Wideröe[3] obtained a doubling of the energy of particles, while in 1931 Sloan and Lawrence[9] were able to obtain a 30-fold multiplication of voltage and produced 10^{-7} amperes of 1,260,000 volt singly charged mercury ions.

An important development occurred in 1932 when Lawrence and Livingston[5] suggested the use of a magnetic field to bend the particles round in a circle and thus use the same accelerating gap over and over again. It is possible to do this because with non-relativistic particles, the time taken to move round a semi-circle in a magnetic field is independent of their velocity. This is the arrangement used in the cyclotron which has been developed into such an important tool by Lawrence.

In 1929 it seemed that much development work would have to be done on the indirect methods of obtaining fast particles, while at the same time there were indications that, after all, nuclear disintegrations might be produced by particles of reasonably low energies. The application of the wave mechanics showed that there was a non-zero probability that particles might penetrate barriers which they could not surmount. If a sufficiently great number of low energy particles were directed against the barrier, some would go through. Using these ideas, Gamow[10] was able to explain the Geiger-Nuttall law for the emission of α-particles. In 1928 he visited the Cavendish Laboratory, and Cockcroft discussed with him the reverse problem of getting particles inside a nucleus. As a result of this, Cockcroft sent a memorandum to Rutherford in which he showed that protons of only 100,000 electron-volts energy had a small but not negligible chance of penetrating the nucleus of a light atom. As it was expected that very large numbers of protons could be accelerated by this voltage, an appreciable number of penetrations of the barrier should occur. This result had the effect of encouraging the transfer of attention from the indirect methods, which appeared to be a long-term project, to the production of fast particles by using high potentials of not unreasonable magnitude.

In the work carried out by J. D. Cockcroft[11] and the writer, the aim was to accelerate particles by the direct application of potentials of up to 300 kilovolts, these being about the highest potentials which it had been found possible to apply to a vacuum tube for the production of X-rays and cathode rays. The conventional tubes of the time were large glass bulbs with two stems and these were used both for the rectifiers and for the accelerating tube. They may be seen in the photograph (Fig. 1). The transformer voltage

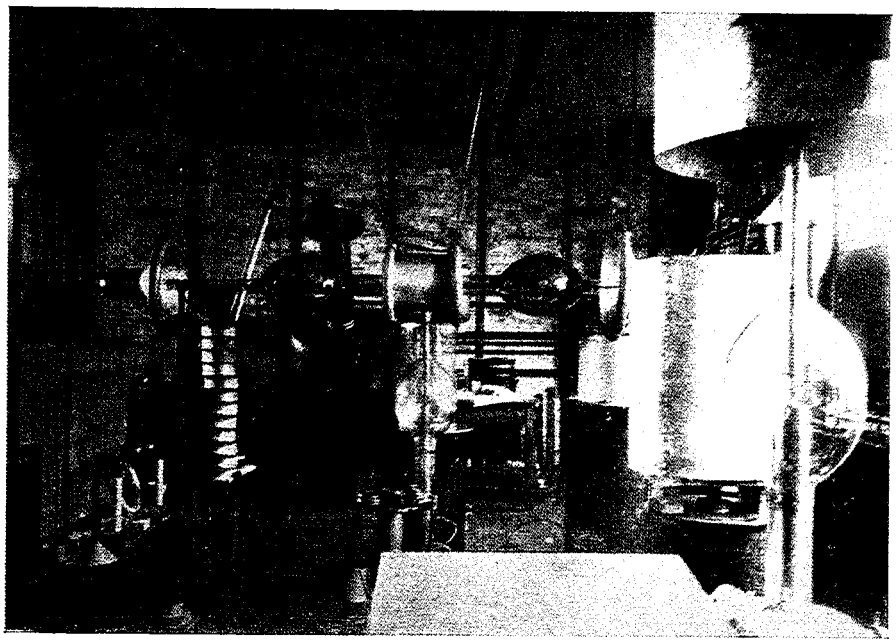


Fig. 1. *By courtesy of The Royal Society.*

was rectified by the two horizontal rectifiers placed in series, these being evacuated through a third bulb connecting them to the pumps. These are shown in the centre of the picture while the accelerating tube is on the right.

With this apparatus several microamperes of protons accelerated by about 280 kilovolts could be obtained.

At this stage the laboratory used had to be vacated and a much larger room became available. Taking advantage of this, the production of much higher energy particles was attempted[12]. For this purpose the voltage multiplier circuit shown in Fig. 2 was used. It is a modification of one due to Schenkel[13]. It gives a fourfold multiplication of voltage and is capable of extension to any even multiple of the transformer voltage. Essentially, it consists of condensers C_1 and C_2 in series, the voltages across them being maintained equal by means of the transfer condenser C_3. This condenser is connected, in effect, alternately in parallel in rapid succession across C_1 and C_2. C_1 becomes charged to twice the peak voltage (V) of the transformer because during one half-cycle, C_4 is charged to V through the rectifier D_1, while during the next half-cycle the voltage across C_4 is added to the trans-

former voltage and thus C_1 gets charged through the rectifier D_2 to twice the transformer voltage.

In addition to giving a steady voltage which may be any desired even multiple of the transformer voltage, the circuit has other advantages. It gives steady voltage tappings at intermediate points, these being useful when using a multi-section accelerating tube. The rectifiers are all connected in series and so may be erected as a single column and evacuated by one diffusion pump at earth potential. They, as well as the accelerating tube, were made out of straight glass cylinders, these being found to withstand high voltages much better than the largest glass bulbs obtainable.

Fig. 3 is a photograph of the high-voltage equipment at the Cavendish Laboratory as it appeared towards the end of 1931. The tower of four rectifiers is on the left and the 2-section accelerating tube is on the right of the centre of the picture. They were evacuated by separate oil diffusion pumps of the type which had recently been developed by C. R. Burch[14] at Metropolitan Vickers. Their use simplified greatly the problem of maintaining a sufficiently low pressure in the apparatus. High-vacuum technique was also made much simpler by the use of « plasticene » and later by the use of Apieson Sealing Compound Q. The pieces to be joined had merely to be placed

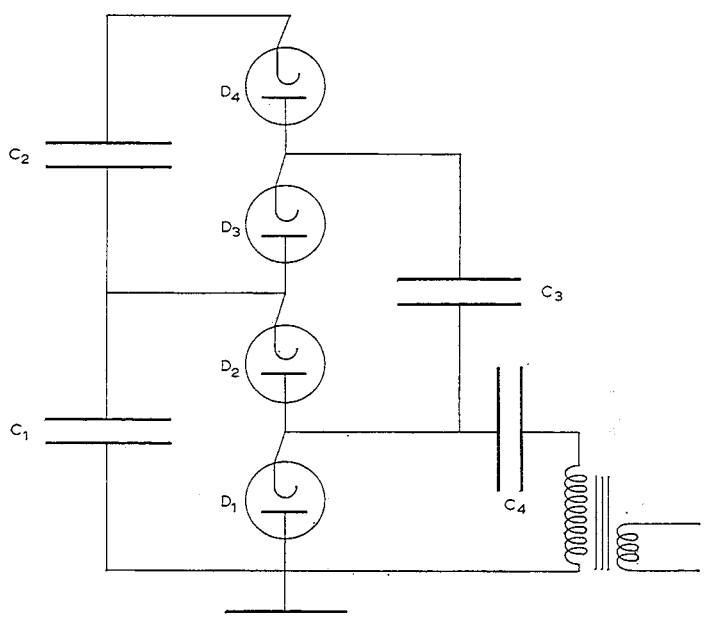
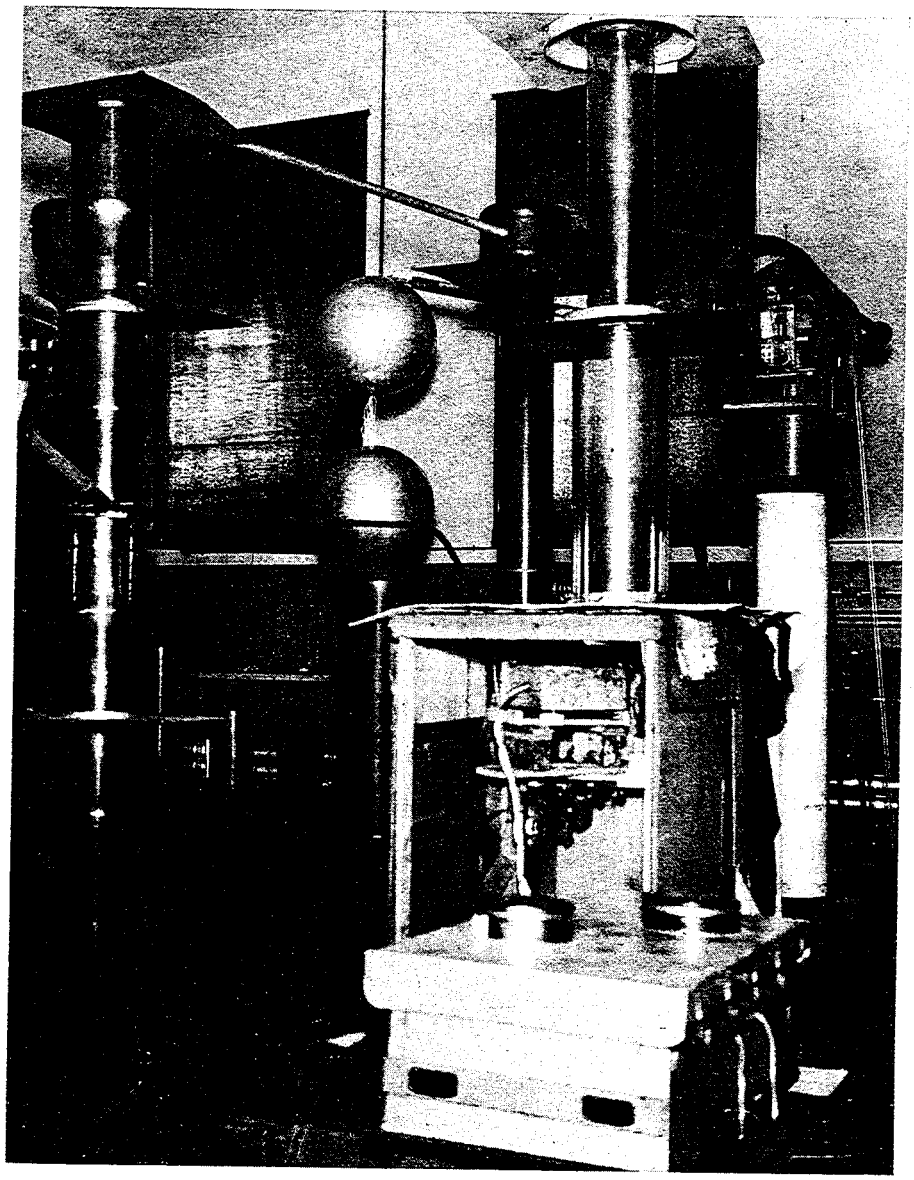


Fig. 2.



By courtesy of the Keystone Press Agency.

Fig. 3.

together and the joint made vacuum tight by pressing the compound with our fingers.

The proton source was of the type used by Aston, suitable power supplies being obtained from the belt-driven equipment on top of the tall white porcelain cylinder. Currents of up to about 10 microamperes accelerated by potentials of up to about 700 kilovolts were obtained. These were used to bombard various targets placed at the bottom of the tube. The disintegrations which were found to be produced are being described by Sir John Cockcroft in his Nobel Lecture.

The discovery that light elements could be disintegrated by artificially accelerated particles gave an additional impetus to development work on the various methods of producing them. The only other direct method which has been found suitable is that of the Van de Graaff machine[15]. In it an electrical charge is sprayed on to a fast-moving belt and is thus carried to the high-voltage electrode. It is suited to nuclear disintegration work because, like all electro-static machines, it readily gives small currents at high voltages. They are usually operated under pressure in a steel cylinder in order to increase both output current and voltage. They appear to be very convenient and economical for voltages in the range from about one to about five million volts.

Later development of the indirect methods. As the size of cyclotrons increases and faster particles are produced, a difficulty arises due to the relativistic increase of mass of the particle. The particle tends to get out of step with the accelerating voltage. This limits the number of times the particles can be allowed to travel round. Hence higher accelerating voltages must be applied to the dees. This means greatly increased high-frequency power and a larger gap in the magnet to give increased clearances. These difficulties can be avoided by the use of an alternating potential whose frequency changes during the course of the acceleration of a particle, but its results in the output current being reduced to a burst of fast particles at the end of each duty cycle of the high-frequency change. Appreciable current can be obtained because of the principle of phase stability (McMillan[16], Veksler[17], Oliphant[18]) which operates when relativistic speeds are reached. The effect produced is that particles, which are tending to get out of step with the accelerating voltage, have their phase automatically changed so as to bring it back to the phase stable point. Hence there is no objection to having the many accelerations necessary when quite low voltages are used on the dees, and no restrictions are placed on the exact way in which the frequency is changed. This is the

method of the frequency-modulated cyclotron or synchro-cyclotron. The largest of these machines gives particles of about 350 MeV energy. To go to higher energies will require the construction of larger magnets than have yet been built and their cost is likely to increase with at least the cube of the pole diameter. A considerable saving in the cost of the magnet can be effected if a particle can be made to move in a circle of constant radius for the whole of its acceleration. This is done in the synchrotron by increasing the magnetic field gradually during the course of the acceleration. The magnet must be laminated and, as its time constant is large, burst of fast particles are obtained only at intervals of the order of 10 seconds. The method promises to be useful in the range 1,000 to 10,000 MeV. The latter figure may well be the economic limit of the method.

A linear accelerator has the advantage that no magnet is required and that its cost should not rise much more steeply than with the energy of the particles required. Recent progress made in microwave technique has made it possible for energies of up to about 30 MeV to be obtained and the method may well prove to be useful for much higher energies.

Looking to the future, it is difficult to see how particles of energy greater than 10,000 MeV can be produced economically by existing methods. Further progress may have to await the introduction of new ideas.

1. E. Rutherford, *Phil. Mag.*, 37 (1919) 581.
2. J. Slepian, *U.S. Patent No. 1, 645, 304* (1927); filed (1922).
3. R. Wideröe, *Arch. Electrotech.*, 21 (1928) 387.
4. E. T. S. Walton, *Proc. Cambridge Phil. Soc.*, 25 (1929) 469.
5. E. O. Lawrence and M. S. Livingston, *Phys. Rev.*, 40 (1932) 19.
6. D. W. Kerst and K. Serber, *Phys. Rev.*, 60 (1941) 53.
7. D. W. Kerst, *Phys. Rev.*, 60 (1941) 47.
8. G. Ising, *Arkiv Mat. Astron. Fysik*, 18, Nr. 30 (1925) 45.
9. D. H. Sloan and E. O. Lawrence, *Phys. Rev.*, 38 (1931) 2021.
10. G. Gamow, *Z. Physik*, 51 (1928) 204; 52 (1928) 51.
11. J. D. Cockcroft and E. T. S. Walton, *Proc. Roy. Soc. London*, A129 (1930) 477.
12. J. D. Cockcroft and E. T. S. Walton, *Proc. Roy. Soc. London*, A136 (1932) 619.
13. M. Schenkel, *Elektrotech. Z.*, 40 (1919) 333.
14. C. R. Burch, *Proc. Roy. Soc. London*, A123 (1929) 271.
15. R. J. van de Graaff, *Phys. Rev.*, 38 (1931) 1919.
16. E. M. McMillan, *Phys. Rev.*, 68 (1945) 143.
17. V. I. Veksler, *J. Phys. U.S.S.R.*, 9 (1945) 153.
18. M. L. E. Oliphant, Proposal submitted to Directorate of Atomic Energy, Department of Scientific and Industrial Research, London, 1943.

Biography

Ernest Thomas Sinton Walton was born at Dungarvan, County Waterford on the south coast of Ireland on October 6th, 1903, the son of a Methodist Minister from County Tipperary. The ministry demanded that his father move from place to place every few years, and he attended day schools in Banbridge (County Down) and Cookstown (County Tyrone). In 1915 he was sent as a boarder to the Methodist College, Belfast, where he excelled in mathematics and science, and in 1922 he entered Trinity College, Dublin, on a scholarship. He read the honours courses in both mathematics and experimental science, specializing in physics, and graduated in 1926 with first-class honours in both subjects; he received his M.Sc. degree in 1927.

In 1927, he was awarded a Research Scholarship by the Royal Commissioners for the Exhibition of 1851 and he went to Cambridge University to work in the Cavendish Laboratory under Lord Rutherford. He continued at Cambridge after receiving a senior research award of the Department of Scientific and Industrial Research in 1930, and received his Ph.D. in 1931. Walton was Clerk Maxwell Scholar from 1932 to 1934 when he returned to Trinity College, Dublin, as Fellow: he was appointed Erasmus Smith's Professor of Natural and Experimental Philosophy in 1946, and in 1960 he was elected Senior Fellow of Trinity College.

Prof. Walton's first researches involved theoretical and experimental studies in hydrodynamics and, at the Cavendish Laboratory, he worked on indirect methods for producing fast particles, working on the linear accelerator and on what was later to become known as the betatron. He followed this with work on the direct method of producing fast particles by the use of high voltages this work being done jointly with J. D Cockcroft. A suitable apparatus was built which made it possible to show that various light elements could be disintegrated by bombardment with fast protons. They were directly responsible for disintegrating the nucleus of the lithium atom by bombardment with accelerated protons, and for identifying the products as helium nuclei.

Prof. Walton has taken part in many activities outside his academic work,

and he has served on committees connected with the Dublin Institute for Advanced Studies, the Institute for Industrial Research and Standards, the Royal City of Dublin Hospital, the Royal Irish Academy, the Royal Dublin Society, Wesley College, Dublin, and many government and church committees. He has had numerous scientific papers published in the journals of learned societies, particularly on the subjects of hydrodynamics, nuclear physics, and microwaves.

He was awarded the Hughes Medal, jointly with Sir John Cockcroft, by the Royal Society of London in 1938, and in 1959 he received an honorary Doctor of Science degree from Queen's University, Belfast.

E. T. S. Walton married Freda Wilson, daughter of a Methodist Minister and a former pupil of Methodist College, Belfast, in 1934. They have two sons and two daughters, Alan, Marian, Philip, and Jean.

Physics 1952

FELIX BLOCH

EDWARD MILLS PURCELL

«for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith»

Physics 1952

Presentation Speech by Professor E. Hulthén, member of the Nobel Committee for Physics

Your Majesty, Your Royal Highnesses, Ladies and Gentlemen.

For the man in the street I suppose the compass needle is the most familiar magnetic instrument. But when and where the compass was first used is a much-debated question, where we grope between Chinese records from the year 2,600 B.C. and ship's logs made by the Norsemen in their Icelandic voyages in the 12th and 13th centuries. It is typical of all such records, whether of gunpowder or the compass, that they refer to inventions that had long been in use. The very idea of invention, of having been the first, had doubtless not the same significance formerly that it has today. As a matter of fact, the scientific study of magnetism in the sense in which we understand it, was begun only with the publication in London of Gilbert's work *De Magnete* in the year 1,600 A.D. The subsequent investigation and classification of magnetic substances led to their division into three categories: the ferromagnetics or strong magnetics such as iron, cobalt and nickel; the paramagnetic or weak magnetics, including chiefly crystals and fluids; and finally the diamagnetics, with their magnetic repulsion, a property intrinsic in all substances. A compass needle made of a diamagnetic substance turns at right angles to the magnetic lines of force, and thus comes to point in an east-westerly direction. Fortunately, diamagnetism is too weak to cause shipwreck in this way. This wealth of magnetics is today joined by a fourth category, the nuclear magnetism deriving from the atomic nucleus.

The magnetic field radiating from the infinitesimally tiny atomic nucleus is so feeble that its existence was still scarcely more than divined only fifteen or twenty years ago. Thus when Bloch and Purcell, this year's Nobel Prize winners in Physics, are able to register nuclear magnetism with a precision exceeding almost all other measurements in physics, one supposes that this must be thanks to the use of special methods and accessories. But what interest or useful purpose may conceivably be served by such subtleties?

If we consider the methods that have been employed, we soon recognize the idea that runs through all more advanced measurements of a body's magnetic moments. Thus the celebrated German mathematician and physicist,

Karl Friedrich Gauss, determined in 1836 the magnetic moment of the compass needle in relation to its moment of inertia, simply by observing the oscillations of the needle in a magnetic field of known strength.

Now the electrons or the atomic nucleus do not, it is true, behave in quite the same way as the compass needle in the magnetic field, but rather in the manner of the top, the gyroscope, which spins and precesses about the perpendicular. But the electronic and nuclear spins are just as characteristic for these particles as are their electric charge and mass (the atomic weights), so that the deep import of a determination of their gyromagnetic indices becomes immediately obvious.

Now what possibilities exist for the observation and measurement of the frequencies of the electrons and the atomic nucleus in the magnetic field? This is where the new phase in the development comes in. In this connection I need only remind you of the resonance between our radio apparatuses and radio waves. The comparison is actually quite justified, as the electronic and atomic nuclear frequencies in the magnetic field fall precisely within the region for the short-wave radio with wavelengths varying between some tens of meters and the centimeterwaves employed in radar technique.

These atomic frequencies in the magnetic field are so characteristic for each element and its isotopes that they are more undisturbed and regular than the balance-wheel, pendulum and vibrating quartz-crystal in our modern chronometers.

The method for the determination of the nuclear magnetic moment through resonance with radio waves has long been well-known, and was rewarded by the Academy of Sciences with the Nobel Prize for the year 1944 to Rabi. It was with similar methods that the paramagnetism of crystals deriving from the electronic spin was investigated by Gorter in Leiden.

Rabi carried out his investigations on nuclear magnetic moments according to the molecular-ray method, an artificial method which has, certainly, the inestimable advantage that the investigated substance is in a state of very high rarefaction, though at the same time this limits its application. The methods of Purcell and Bloch imply a great simplification and generalization in this respect, which enables their application to solid, liquid and gaseous substances. This brings us to the useful purposes which may be served. Since each kind of atom and its isotopes have a sharply defined and characteristic nuclear frequency, we can in any object placed between the poles of an electromagnet seek out and examine with radio waves all the various kinds of atom and isotopes present in the object in question, and, *which is the essen-*

tial point, this without in any perceptible way affecting the same, its form, crystalline structure, etc. This form of analysis *in situ* is therefore probably not parallelled in any other known methods of analysis. Its extraordinary sensitiveness also makes it particularly well-adapted as a micro-method in many scientific and technical fields.

Professor Purcell. As far I have been able to follow your activities since you stopped working at the great Radiation Laboratory at M.I.T. at the end of the War, and up to your development of the excellent method of *nuclear resonance absorption* for which you have been awarded your Nobel Prize, you have happily realized man's old dream of beating the sword into a ploughshare. Your wide experience in electronics and the deep interest you early showed in paramagnetic phenomena may thus conceivably have contributed to the invention of your method, which through its extraordinary sensitiveness gives us a deep insight into the constitution of crystals and fluids, and the interactions, so-called relaxations, between the tiniest particles of matter.

In part with this method, and in part without it, you and your collaborators have made a number of important discoveries, among which I would like particulary to stress the three following:

Your method for studying nuclear magnetic resonance in weak magnetic field produced according to the solenoid method, which is of great value for the absolute determination of nuclear magnetic moments.

In the very interesting experiment which you performed together with Dr. Pound, you have produced with paramagnetic resonance the rather unique situation in which the state of the atomic nucleus corresponds to negative temperatures in the absolute-temperature scale.

Finally, as a quite spectacular discovery I may mention your observation with Dr. Ewen in 1951 of a line in the galactic radiospectrum caused by atomic hydrogen, an important contribution to radioastronomy.

Please accept our congratulations, and receive your Nobel Prize from the hands of His Majesty.

Professor Bloch. It would be difficult in the few minutes at my disposal to try to give the main features of the nuclear induction method for which you have been awarded your Nobel Prize. It would be still more difficult for me to give an exhaustive account of the ways that led you to this invention.

You began your career as a theoretical physicist, well-known for your fundamental contributions to the theory of metals.

When, quite unexpectedly, you went over to experimental research, this must have been, I feel, with deliberation and assurance. For you had in your kitbag a tool of extraordinary value, the method for the magnetic polarization of a beam of neutrons. The inestimable value of possessing a good idea, of indefatigably testing and perfecting it, is best illustrated by your precision-measurements of the magnetic moment of the neutron, one of the most difficult and at the same time most important tasks in nuclear physics.

But ideas give birth to new ideas, and it was, as I understand, in this way that you hit upon the excellent notion of eliminating the difficult absolute determination of the magnetic field by a direct measurement of the neutron moment in units of the proton cycle (the nuclear magneton). According to your own account it was this solution which finally led you to the nuclear induction method.

In congratulating you I now beg you to receive your Nobel Prize from the hands of His Majesty.

FELIX BLOCH

The principle of nuclear induction

Nobel Lecture, December 11, 1952

It is a tribute to the inherent harmony and the organic growth of our branch of science that every advance in physics is largely due to the developments that preceded it. The discovery for which Purcell and I have received the honor of the Nobel Prize award for the year 1952 is a typical example of this situation, and before describing the principle I shall therefore present an outline of its long and distinguished background.

Both the method and the object go back ultimately to spectroscopy, a field to which modern physics owes so much in other respects. Among the various aspects of this field there are two which are of particular importance here: the Zeeman effect for introducing magnetic fields as an essential element of spectroscopy, and the hyperfine structure of spectral lines for revealing the existence of nuclear moments. The correct interpretation of hyperfine structures was first given in 1924 by Pauli[1], who proposed that atomic nuclei may possess an intrinsic angular momentum (spin) and, parallel to its orientation, a magnetic moment. The energy of interaction of this magnetic moment with the magnetic field $H_{(0)}$, produced by the atomic electrons at the position of the nucleus, depends upon the angle between them and leads thus to the observed small splitting of the energy levels. Conversely, it is possible under suitable conditions to determine from this splitting both the spin and the magnetic moment of the nucleus, and these two important quantities have indeed been determined in a great number of cases from the observation of hyperfine structures. The magnetic moments of the nuclei have been found, in all observed cases, to be of the order of the «nuclear magneton» which one obtains by substituting in the formula for the atomic Bohr magneton the mass of the proton in place of that of the electron. Nuclear moments are thus about a thousand times smaller than atomic moments, and this is plausible in view of the fact that one deals here with protons instead of electrons as elementary charged constituents. There are, however, distinct disadvantages in the optical determination of nuclear moments. In the first place the accuracy is seriously limited due to the fact that the effect consists only in such a small splitting of spectral lines that one has to

be concerned with their finite width. In the second place it is necessary for the determination of the nuclear magnetic moment from the observed hyperfine structure to have a knowledge of the field $H_{(o)}$ which is usually rather inaccurate since it involves complex electron configurations. In view of these limitations one is led to values of nuclear magnetic moments with an accuracy of a few percent at best. Finally, one is faced with the fact that hyperfine splittings tend to decrease with decreasing atomic number with the result that it is not possible, by optical means, to observe them in the case of the greatest fundamental importance, that of hydrogen.

A decisive step forward was made in 1933 by Stern[2], who applied his method of molecular beams to the determination of the magnetic moments of the proton and the deuteron in hydrogen molecules. Instead of the emitted light, it is here the deflection of the molecule in an inhomogeneous magnetic field which is affected by the nuclear moments. Although the observed effect was close to the limit of observability it yielded the proton moment to within ten percent with the most important result that instead of having the expected value of one nuclear magneton it is about 2.5 times larger. Of similar importance was the result that the magnetic moment of the deuteron was between 0.5 and 1 nuclear magneton, since it indicated from the simplest plausible considerations of the structure of this nucleus that one should ascribe a moment of about 2 nuclear magnetons to the neutron. I shall come back later to this point; it represents the start from which my own experimental work has followed in an almost continuous line.

Subsequent to Stern's work a number of far-reaching further developments have been achieved by Rabi in the application of atomic and molecular beams to the measurement of nuclear moments and hyperfine structures. Without attempting completeness I want to mention some aspects of method in the brilliant series of investigations which he carried out with his collaborators. One of them is based upon a paper by Breit and Rabi[3], which treats the variation of the magnetic moment of an atom for the different Zeeman levels of hyperfine structure under the influence of an external magnetic field and which was applied to atomic beams where the deflection gives a direct measure of the magnetic moment. Another important aspect lies in the passage of the beam through two and later three separate field regions which can be adjusted to give zero deflection so that one deals with a null method. These innovations, besides giving many other interesting results, allowed the measurement of the hyperfine structure in the ground state of light and heavy hydrogen atoms; since the previously mentioned field $H_{(o)}$, produced

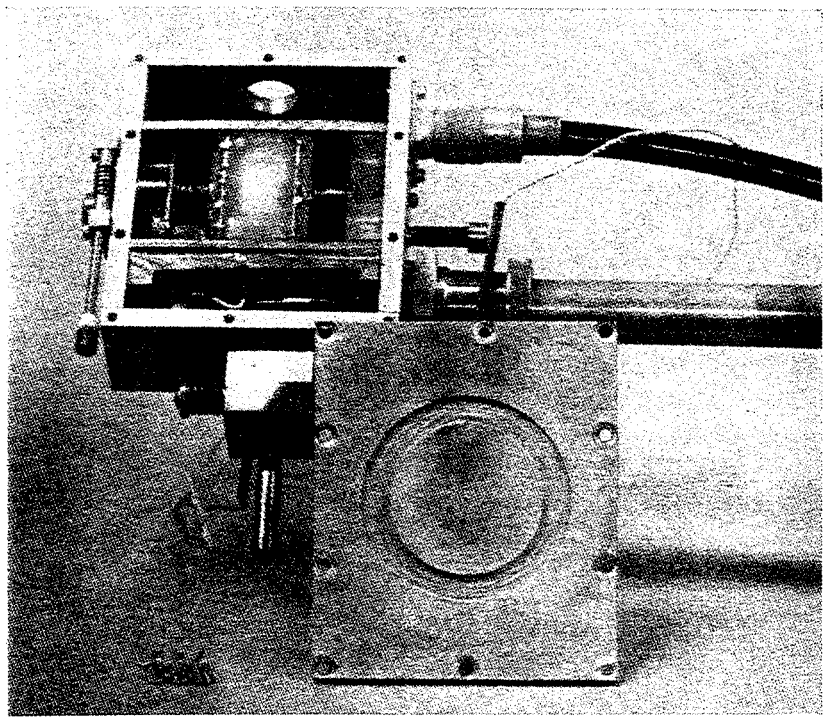


Fig. 1. The «head» of the crossed-coil arrangement with the cover plate removed. The bottom tube to the right contains the leads to the transmitter coil, which is wound in two sections visible in black in the head. The black cable leads from the receiver coil to the amplifier; the receiver coil is wound with a vertical axis inside the hollow lucite piece between the two sections of the transmitter coil. The sample test tubes are placed in its interior through the circular hole at the top of the supporting frame.

by the electron at the place of the nucleus, was given here, through a formula of Fermi[4], from the well-known theory of the hydrogen atom, this measurement resulted in the determination of the magnetic moments of the proton and the deuteron with an accuracy of a few percent.

However, the most significant improvement in molecular and atomic beam techniques was the introduction of the magnetic resonance method. The beam here passes through a region where the magnetic field is homogeneous and constant with a weak alternating magnetic field superimposed at right angles to the strong constant field. Analogous to the resonance absorption of visible light, transitions occur here from one Zeeman level to another if the alternating field satisfies Bohr's frequency condition for the energy

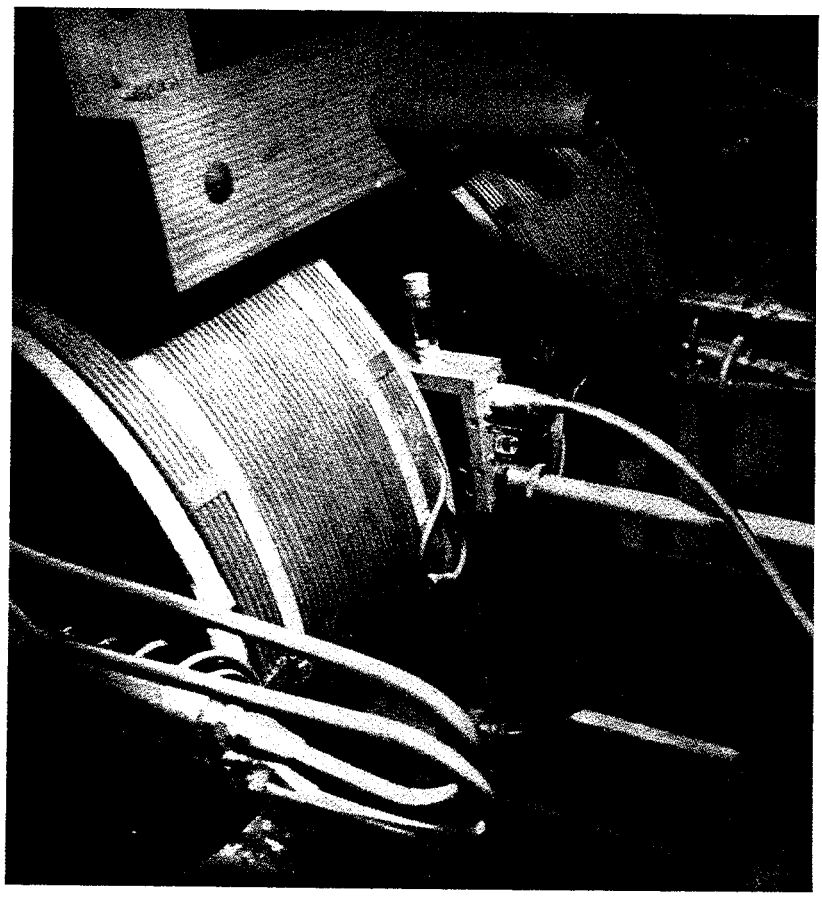


Fig. 2. The same head as in Fig. 1, about to be inserted in the gap of an electromagnet and containing a sample test tube. The two protruding lucite rods between the leads reach into the interior of the head and carry small copper disks; a fine adjustment of the coupling is achieved by rotation of these «paddles».

difference between the two levels. However, instead of optical frequencies one deals here normally with frequencies in the radio range so that this application of the magnetic resonance method, like our own, is properly labelled as belonging to the new field of radiofrequency spectroscopy. In the beam technique it has the great advantage of dispensing with a knowledge of the deflecting inhomogeneous fields, since the deflection is merely used now as an indicator for the occurrence of transitions in the homogeneous field region. A very much greater accuracy can thus be obtained; it led, for ex-

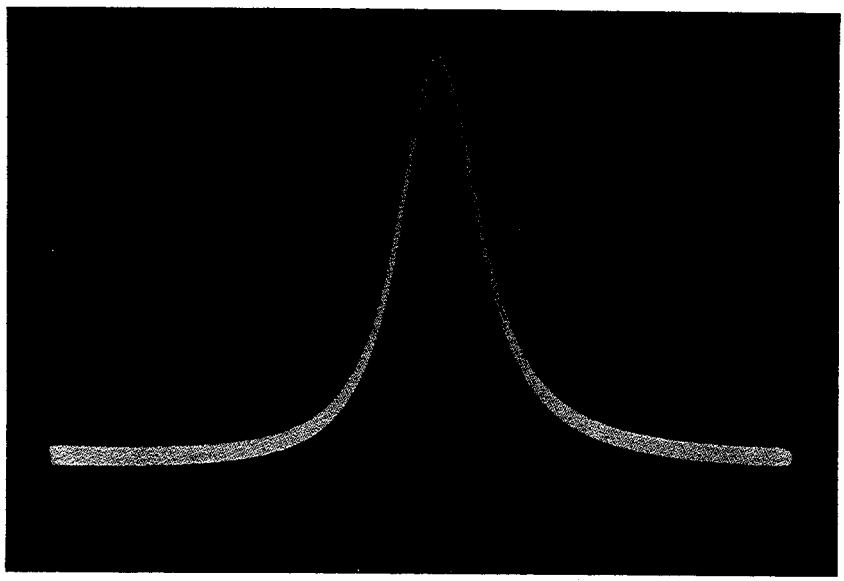


Fig. 3. A resonance line of protons in water, containing $MnSO_4$ as a paramagnetic catalyst and obtained from the phase component of the nuclear induction signal which corresponds to absorption. The photograph is that of the trace on a cathode-ray oscillograph with the *vertical deflection* arising from the rectified and amplified signal, and the *horizontal deflection* corresponding to different values of the constant field.

ample, to the knowledge of the magnetic moments of the proton and the deuteron with an accuracy of about one part in a thousand and to the important discovery of a small but finite electrical quadrupole moment of the deuteron[5] in 1939.

The first use of the magnetic resonance method was suggested in 1936 by Gorter[6] in an attempt to detect the resonance absorption of radio quanta through the heating of a crystal. While the results of this experiment were negative, Rabi[7] in 1937 has treated the transitions in a rotating field and has pointed out their use in atomic and molecular beams.

Coming from quite a different direction, I was led at that time to similar ideas. They originated from my preceding work which dealt with the magnetic moment of the neutron and which had been stimulated by Stern's previously mentioned measurement of the magnetic moment of the deuteron[2]. The idea that a neutral elementary particle should possess an intrinsic magnetic moment had a particular fascination to me, since it was in such striking contrast to the then only existing theory of an intrinsic moment

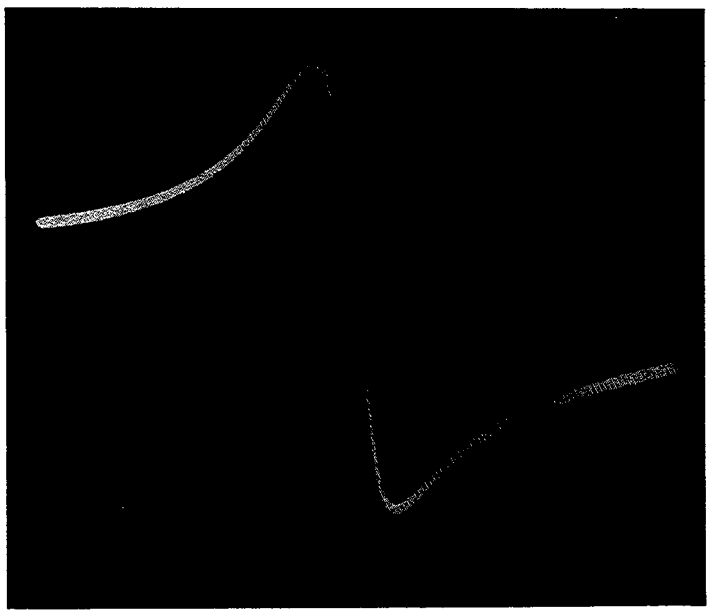


Fig. 4. The only difference between this line and that of Fig. 3 lies in the adjustment of the observed phase which is here that corresponding to dispersion.

which had been given by Dirac[8] for the electron. Combining relativistic and quantum effects, he had shown that the magnetic moment of the electron was a direct consequence of its charge and it was clear that the magnetic moment of the neutron would have to have an entirely different origin. It seemed important to furnish a direct experimental proof for the existence of a magnetic moment of the free neutron, and I pointed out in 1936[9] that such a proof could be obtained by observing the scattering of slow neutrons in iron. The very strongly inhomogeneous magnetic field in the neighborhood of each iron atom was shown to affect a passing neutron through its magnetic moment and thus to lead to an appreciable magnetic scattering effect; it was shown at the same time, that magnetic scattering would lead to the polarization of neutron beams. The existence of this effect was first clearly demonstrated in 1937 by a group of investigators at Columbia University[10], and it opened up the possibility of further work with polarized neutron beams.

The most desirable goal to be reached here was that of accurately measuring the magnetic moment of the neutron. It occurred to me that resonance depolarization could be achieved by passing a polarized neutron beam

through a region where a weak oscillating field is superimposed on a strong constant field, provided that the frequency of the former is equal to the frequency with which the neutron moment carries out a precessional motion around the direction of the constant field. A knowledge of this field and of the corresponding resonance frequency directly determines the magnetic moment under the safe assumption that the spin of the neutron is 1/2, and the magnetic scattering effect enters in this arrangement merely as an indicator for the occurrence of resonance depolarization. The application to polarized neutron beams was also noted by Rabi[7] in his previously mentioned original paper on the magnetic resonance method. It was first achieved in 1939 by Alvarez and myself[11] with the use of the Berkeley cyclotron, and yielded a value for the magnetic moment of the neutron which was consistent with that of the deuteron if one assumed the latter to be additively composed of the moments of the proton and the neutron. The accuracy of this measurement amounted to about one percent and was partly limited by that with which the strength of the constant field could be determined. Another limit of accuracy arose from the smallness of the observed polarization effect, but a subsequent systematic investigation of neutron polarization[12], carried out with the Stanford cyclotron, showed how this effect could be greatly increased.

It was of considerable importance to improve the accuracy of the determination of the neutron moment to at least one part in a thousand in order to test small deviations from the additivity of the moments of the proton and the neutron, which could be expected in connection with the finite electric quadrupole moment of the deuteron, according to the theoretical work of Rarita and Schwinger[13]. The fact that higher accuracy hinged essentially upon that of a field calibration and the search for a suitable and convenient standard led me to new ideas when, toward the end of the last War, my thoughts turned back to the continuation of my previous work.

The essential fact of the magnetic resonance consists in the change of orientation of nuclear moments, and the methods to be employed in molecular and atomic beams as well as in neutron beams are primarily indicated by the different ways to detect this change. The acquaintance with radio techniques during the War suggested to me still another and much simpler way, that of detecting the reorientation of nuclear moments through the normal methods of radio reception. The signals to be detected would be due to the electromagnetic induction caused by nuclear reorientation and should appear as a voltage difference between the terminals of an external electric circuit.

I believe that this is the most general and distinctive feature of our discovery, and it is for this reason that I chose for it the name of « nuclear induction ». Purcell, whose independent approach was largely based on considerations of energy relations, has chosen to call it « nuclear magnetic resonance absorption », but soon after our respective initial work and despite its apparent difference, it became clear that it was based upon the same principle.

In order to understand this principle, one can start from macroscopic quantities and describe the underlying phenomenon in classical terms. Consider for this purpose, as a typical example, about one cubic centimeter of water with the protons contained in it as the nuclei under investigation. Their magnetic moments are oriented in a completely random manner in the absence of an external magnetic field; after the sample has been brought into such a field, however, there will be established a new thermal equilibrium in which the magnetic moments are distributed with a slight surplus parallel to the field. Even in relatively strong fields of the order of 10,000 gauss this surplus will at room temperature amount to no more than about one part in a million. While its direct observation would be difficult, there thus exists a « nuclear paramagnetism » in the sense that one deals with a finite macroscopic nuclear polarization which is both parallel and proportional to the applied external field. The establishment of thermal equilibrium demands the transfer of the energy released by the partial orientation of the nuclear moments into heat, and it can take place only through interaction of these moments with their molecular surroundings. The strength of this interaction determines the time interval required for the nuclear moments to adjust themselves to the equilibrium conditions; it is measured by the « relaxation time », as in the analogous case of atomic paramagnetism. The role of the relaxation time is of basic significance for our experiments, and I shall soon come back to its discussion.

For the moment, we shall return to the equilibrium state, once it is established, and to the description of the nuclear polarization under the conditions of magnetic resonance. A simple mechanical consideration of the gyroscope shows that an alternating field at right angles to the constant field has the effect of tilting the direction of the polarization with respect to the constant field and that the polarization will thereupon continue to perform a precessional rotation around this field. The angular frequency of precession is proportional to the field with a constant of proportionality which is called the « gyromagnetic ratio » of the nuclei and which is equal to the ratio of their magnetic moment and their intrinsic angular momentum. From a perfectly

macroscopic point of view, one thus deals with a situation in which the protons in our cubic centimeter of water have the effect of an invisible compass needle rotating in its interior. The «invisibility» refers actually only to observation of optical frequencies; the rotation occurs in the range of radio-frequencies, and it can very well be observed by using Faraday's law of induction. Indeed, the rotation of our compass needle is accompanied by that of a magnetic field which possesses an alternating component perpendicular to the axis of rotation, and hence by an electromotive force, induced in a suitably wound coil of wire around the sample. From here on it is merely a matter of the standard techniques of radio reception to rectify and amplify this electromotive force so that it can be recorded on a volt-meter, displayed on a cathode-ray oscillograph, or made audible in a loudspeaker.

What amazed me most in my first calculations on this effect was the magnitude of the signals which one could expect from nuclear induction. In our example of a cubic centimeter of water in a normal field of a few thousand gauss they turned out to amount to the order of a millivolt. This magnitude is well above the noise which accompanies any radio receiver and which sets the ultimate limit of signal detection. It should be observed here that, being a phenomenon of fluctuations, the noise can always be reduced by averaging over sufficiently long times. This procedure was used later to very greatly increase the sensitivity of the method; it is characteristic of the present possibilities that my collaborators have succeeded in the last few years in detecting in natural water signals arising from deuterium and from the isotope of oxygen with atomic mass 17, despite their low abundances of 0.02 and 0.04 percent, respectively.

The existence and detection of a precessing nuclear polarization in a sample represents to my mind the basis of nuclear induction. It is, however, necessary to consider also the features which produce and counteract the tilt of the polarization with respect to the constant field. Magnetic resonance enters here as the most important means of producing the tilt, since it allows its achievement under the application of relatively weak oscillating fields. In fact, it is a common feature of every resonance phenomenon that relatively weak external forces can produce large effects if their frequency is equal to the natural frequency of the system to which they are applied. The natural frequency in question is, in our case, that with which the nuclear polarization precesses by itself around the constant field and the practical way to determine this frequency is to vary either that of the applied alternating field or the magnitude of the constant field until resonance conditions are

established and detected by a maximum of the observed nuclear induction signal. The simultaneous knowledge of resonance field and frequency then directly yields, as in the use of magnetic resonance in molecular beams, the gyromagnetic ratio and, with a knowledge of the spin, the magnetic moment of the nucleus. Actually, it is also possible to determine the spin separately by using the additional piece of information contained in the intensity of the observed signal.

To follow the analogue of mechanical resonance we must now come back to relaxation, which can be seen to act like a friction, and which counteracts the tilt produced by the alternating field. If the friction is large, i.e., if the relaxation time is short, it will either reduce the effect for a given amplitude or require a correspondingly larger amplitude of the alternating field. It will, in either case, result in a relatively broad resonance line, thus diminishing the accuracy of the measurement. While from this point of view it is undesirable to have too short a relaxation time, it is equally undesirable to have it too long, since the very circumstance of producing its tilt diminishes the magnitude of the polarization so that it requires the refreshing influence of the relaxation mechanism to bring it back to its equilibrium value.

There was not much known about the magnitude of nuclear relaxation times when Purcell and I started our first experiments on nuclear induction, and the main doubt about their success arose from the possibility of insufficient relaxation. In fact, it seems, in retrospect, that the failure of Gorter's first attempt[6], as well as of a second one, undertaken in 1942[14], was primarily due to this circumstance. While E. M. Purcell, H. C. Torrey and R. V. Pound[15], toward the end of 1945, obtained their first positive results from protons in paraffin, the late W. W. Hansen, M. E. Packard, and I[16] found ours a few weeks later in water without either of the groups knowing anything about the work of the other. The relaxation time of paraffin has the convenient value of about $\frac{1}{100}$ second, while pure water has a somewhat unfavorably long relaxation time of about 2 seconds. Neither of these two values had been foreseen, and I was fully prepared to find the relaxation time of pure water considerably longer and in fact too long for our method of observation. It was known, however, that the conversion of ortho- and parahydrogen was accelerated by the presence of paramagnetic atoms and molecules; this mechanism has the common feature, with the attainment of the equilibrium polarization of protons, that it requires a random process of nuclear reorientation, and it had been understood to take place through the magnetic field of the paramagnetic catalyst acting upon the magnetic mo-

ment of the proton. An estimate showed that, depending upon the concentration of a paramagnetic salt dissolved in water, a wide range of relaxation times, going down to values of the order of 10^{-5} second, could be obtained. Before starting our observations we therefore had prepared solutions of the paramagnetic iron nitrate in water, and although the first weak signals were received from pure water we found, shortly afterward, considerably stronger signals from a solution with about one-half molar concentration of iron nitrate. The signals appeared as rather broad lines on the cathode-ray oscillograph because of insufficient homogeneity of the constant magnetic field.

Since the width of the resonance line determines the accuracy with which magnetic moments can be determined, we shall briefly consider the conditions necessary for obtaining sharp lines. In the first place, it is necessary that the constant field have the same value in all parts of the sample; in the second place, one must not choose an excessive amplitude of the alternating field, since this too would cause an excessive broadening. The ultimate limit is given by the natural width of the line and it is closely related to the relaxation time; it can be seen, in fact, that the relative accuracy of a measurement by nuclear induction, due to the natural line width, is limited to the order of the number of cycles which the nuclear polarization carries out in its precession around the constant field during the relaxation time. As an example, we shall again consider protons in pure water and in a field of 10,000 gauss; the frequency of precession is here 42.5 megacycles per second, so that about 10^8 cycles are performed during the relaxation time of approximately 2 seconds. This means that an accuracy of about 1 part in 100 millions could be, in principle, achieved here, provided that one had a sufficiently homogeneous field available. While this limit has not yet been reached, it is noteworthy that in water, alcohol, and other liquids, resolutions of one part in 10 millions have actually been achieved. It is indeed the possibility of coherent observation over a large number of cycles which allows the use of nuclear induction as a method of high precision measurements. In fulfillment of my original plans, it was applied by H. H. Staub, D. B. Nicodemus and me to the magnetic moments of the proton, the neutron and the deuteron[17], and it resulted not only in the verification of the previously mentioned deviation from additivity in the deuteron[13], but in its measurement with an accuracy which is beyond the present scope of the theory of nuclear forces. It was particularly gratifying to me to obtain these results from experiments combining the polarization and magnetic resonance depolarization of neutrons with nuclear induction.

The description of nuclear induction which I have presented follows closely my own original thoughts on the subject but it can equally well be approached from other angles. The simplest one is probably that of Gorter[6] in his first attempt to detect nuclear magnetic resonance. We have seen before that the alternating field tilts the nuclear polarization against the constant field. This process requires a certain amount of work which, through relaxation, will reappear in the form of heat produced in the sample. The effect in fact does not involve induction but represents pure nuclear resonance absorption; however, it would be very slight and has not yet been established. A second attempt of Gorter[14], carried out later, is based upon the fact that the nuclear paramagnetic susceptibility has a maximum for radiofrequencies, corresponding to magnetic resonance conditions; it would manifest itself in the frequency of an electric oscillator of which a coil, surrounding the sample, forms the self-inductance. This scheme is actually one of the many others which can be devised for the observation of nuclear induction and, if successful, would have represented the first demonstration of the effect. Purcell's first successful experiment involved the electrodynamical aspect of absorption insofar as its occurrence under resonance conditions was manifested through the increased loss of a cavity resonator; the cavity was replaced in his succeeding arrangements by more conventional circuit elements. A particularly suitable and convenient arrangement consists of a radiofrequency bridge, which contains in one arm a coil, surrounding the sample. As a consequence of nuclear induction there occurs, under resonance conditions, a change of the impedance of this coil and thereby a readily detectable change in the balance of the bridge. It should be remarked that the change of impedance is complex, with its real part corresponding to absorption, its imaginary part to dispersion. This fact can be traced back to the phase relation between the nuclear induction signal and the applied radiofrequency field, and the phase sensitivity of the bridge allows the observation of the effect either as absorption or as dispersion or as a combination of both.

Finally, I shall give a brief description of our own original arrangement which we still use in its principal features. The essential balance which Purcell has obtained by a bridge method is here to a large extent achieved geometrically, by using two radiofrequency coils with their axes oriented at right angles to each other and to the constant field. One of them, the « transmitter coil » produces the alternating field, while the other, the « receiver coil », serves for detection of the nuclear induction signal (see Figs. 1 and 2). A small amount of coupling between the two coils is admitted to produce

a voltage across the receiver coil, and its phase with respect to the superimposed voltage induced by the nuclei can be adjusted for the observation of either absorption or dispersion in similarity to the bridge method (see Figs. 3 and 4).

A considerable variety of other circuits has been introduced by different investigators. Except for the greater or lesser ease of avoiding instrumental difficulties, they lead to the same ultimate sensitivity and accuracy of the method, since they all observe the same basic phenomenon.

There is, however, one distinctive feature in the crossed-coil arrangement, which automatically yields another significant piece of information. The two coils imply a sense of rotation around the constant field; depending upon whether the nuclear polarization precesses in the same or the opposite sense of rotation there results a phase difference of 180 degrees between the voltage in the receiver coil due to coupling with the transmitter coil and the superimposed voltage due to nuclear induction. The action of the rectifier translates this phase difference into an inversion of the signal, which is directly displayed on the oscillograph or on the recording instrument. One obtains in this simple manner information about the sign of the magnetic moment of the nucleus, defined by its relative orientation to the angular momentum, since it is this sign which determines the sense of rotation of the nuclear polarization in a given field. The sign of nuclear moments represents an important clue to their interpretation in terms of nuclear structures; usually it is referred to the sign of the proton moment, which has been known for a considerable time to be positive. It has been determined in this manner for a number of nuclei where it was not previously known.

1. W. Pauli, *Naturwiss.*, 12 (1924) 741.
2. R. Frisch and O. Stern, *Z. Physik*, 85 (1933) 4.
   I. Estermann and O. Stern, *Z. Physik*, 85 (1933) 17.
   I. Estermann and O. Stern, *Phys. Rev.*, 46 (1934) 665.
3. G. Breit and I. I. Rabi, *Phys. Rev.*, 38 (1931) 2082.
4. E. Fermi, *Z. Physik*, 60 (1930) 320.
5. J. M. B. Kellogg, I. I. Rabi, N. F. Ramsey, and J. R. Zacharias, *Phys. Rev.*, 55 (1939) 318; 57 (1940) 677.
6. C. J. Gorter, *Physica*, 3 (1936) 995.
7. I. I. Rabi, *Phys. Rev.*, 51 (1937) 652.
8. P. A. M. Dirac, *Proc. Roy. Soc. London*, 117 (1928) 610.

9. F. Bloch, *Phys. Rev.*, 50 (1936) 259; 51 (1937) 994.
10. P. N. Powers, H. G. Beyer, and J. R. Dunning, *Phys. Rev.*, 51 (1937) 371.
11. L. W. Alvarez and F. Bloch, *Phys. Rev.*, 57 (1940) 111.
12. F. Bloch, M. Hamermesh, and H. Staub, *Phys. Rev.*, 64 (1943) 47.
13. W. Rarita and J. Schwinger, *Phys. Rev.*, 59 (1941) 436.
14. C. J. Gorter and L. J. F. Broer, *Physica*, 9 (1942) 591.
15. E. M. Purcell, H. C. Torrey, and R. V. Pound, *Phys. Rev.*, 69 (1946) 37.
N. Bloembergen, E. M. Purcell, and R. V. Pound, *Phys. Rev.*, 73 (1948) 679.
16. F. Bloch, W. W. Hansen, and M. Packard, *Phys. Rev.*, 69 (1946) 127.
F. Bloch, *Phys. Rev.*, 70 (1946) 460.
17. F. Bloch, D. Nicodemus, and H. Staub, *Phys. Rev.*, 74 (1948) 1025.

Biography

Felix Bloch was born in Zurich, Switzerland, on October 23, 1905, as the son of Gustav Bloch and Agnes Bloch (*née* Mayer). From 1912 to 1918 he attended the public primary school and subsequently the « Gymnasium » of the Canton of Zurich, which he left in the fall of 1924 after having passed the « Matura », i.e. the final examination which entitled him to attend an institution of higher learning.

Planning originally to become an engineer, he entered directly the Federal Institute of Technology (Eidgenössische Technische Hochschule) in Zurich. After one year's study of engineering he decided instead to study physics, and changed therefore over to the Division of Mathematics and Physics at the same institution. During the following two years he attended, among others, courses given by Debye, Scherrer, Weyl, as well as Schrödinger, who taught at the same time at the University of Zurich and through whom he became acquainted, toward the end of this period, with the new wave mechanics. Bloch's interests had by that time turned toward theoretical physics. After Schrödinger left Zurich in the fall of 1927 he continued his studies with Heisenberg at the University of Leipzig, where he received his degree of Doctor of Philosophy in the summer of 1928 with a dissertation dealing with the quantum mechanics of electrons in crystals and developing the theory of metallic conduction. Various assistantships and fellowships, held in the following years, gave him the opportunity to work with Pauli, Kramers, Heisenberg, Bohr, and Fermi, and to further theoretical studies of the solid state as well as of the stopping power of charged particles.

Upon Hitler's ascent to power, Bloch left Germany in the spring of 1933, and a year later he accepted a position which was offered to him at Stanford University. The new environment in which he found himself in the United States helped toward the maturing of the wish he had had for some time to undertake also experimental research. Working with a very simple neutron source, it occurred to him that a direct proof for the magnetic moment of the free neutrons could be obtained through the observation of scattering in iron. In 1936, he published a paper in which the details of the phenomenon

were worked out and in which it was pointed out that it would lead to the production and observation of polarized neutron beams. The further development of these ideas led him in 1939 to an experiment, carried out in collaboration with L. W. Alvarez at the Berkeley cyclotron, in which the magnetic moment of the neutron was determined with an accuracy of about one percent.

During the war years Dr. Bloch was also engaged in the early stages of the work on atomic energy at Stanford University and Los Alamos and later in counter-measures against radar at Harvard University. Through this latter work he became acquainted with the modern developments of electronics which, toward the end of the war, suggested to him, in conjunction with his earlier work on the magnetic moment of the neutron, a new approach toward the investigation of nuclear moments.

These investigations were begun immediately after his return to Stanford in the fall of 1945 and resulted shortly afterward in collaboration with W. W. Hansen and M. E. Packard in the new method of nuclear induction, a purely electromagnetic procedure for the study of nuclear moments in solids, liquids, or gases. A few weeks after the first successful experiments he received the news of the same discovery having been made independently and simultaneously by E. M. Purcell and his collaborators at Harvard.

Most of Bloch's work in the subsequent years has been devoted to investigations with the use of this new method. In particular, he was able, by combining it with the essential elements of his earlier work on the magnetic moment of the neutron, to remeasure this important quantity with great accuracy in collaboration with D. Nicodemus and H. H. Staub (1948). His more recent theoretical work has dealt primarily with problems which have arisen in conjunction with experiments carried out in his laboratory.

In 1954, Bloch took a leave of absence to serve for one year as the first Director General of CERN in Geneva. After his return to Stanford University he continued his investigations on nuclear magnetism, particularly in regard to the theory of relaxation. In view of new developments, a major part of his recent work deals with the theory of superconductivity and of other phenomena at low temperatures.

In 1961, he received an endowed Chair by his appointment as Max Stein Professor of Physics at Stanford University.

Prof. Bloch married in 1940 Dr. Lore Misch, a refugee from Germany and herself a physicist.

EDWARD M. PURCELL

Research in nuclear magnetism

Nobel Lecture, December 11, 1952

Professor Bloch has told you how one can detect the precession of the magnetic nuclei in a drop of water. Commonplace as such experiments have become in our laboratories, I have not yet lost a feeling of wonder, and of delight, that this delicate motion should reside in all the ordinary things around us, revealing itself only to him who looks for it. I remember, in the winter of our first experiments, just seven years ago, looking on snow with new eyes. There the snow lay around my doorstep – great heaps of protons quietly precessing in the earth's magnetic field. To see the world for a moment as something rich and strange is the private reward of many a discovery. But I am afraid it has little bearing on the sober question we must, as physicists, ask ourselves: What can we learn from all this about the structure of matter? It is my privilege to tell you now of some of the things that can be learned.

Let us begin with the most direct application of nuclear induction methods, the measurement of nuclear magnetic moments. The basis for this is the resonance condition

$$f = \frac{\mu H_0}{Ih}$$

in which f is the frequency of precession of the axis of nuclear spin in a magnetic field of strength H_0, and μ is the magnetic moment of the nucleus. The number I is the nuclear spin quantum number, an integer or half-integer, and h is Planck's constant. Now H_0, except for a certain slight correction, is simply the field of the magnet in which the substance has been put, and it can be measured. The frequency of precession, once it is detected, is easily measured with high accuracy, and thus one can determine the quantity μ/Ih. However, for practical reasons, it is hard to measure the strength of a magnetic field very precisely. This difficulty is avoided if one is content to measure the *ratio* of the magnetic moments for two different nuclear species. We could, for example, compare the precession frequencies f_H and f_D for protons and deuterons exposed to the *same* magnetic field H_0. A mixture of light

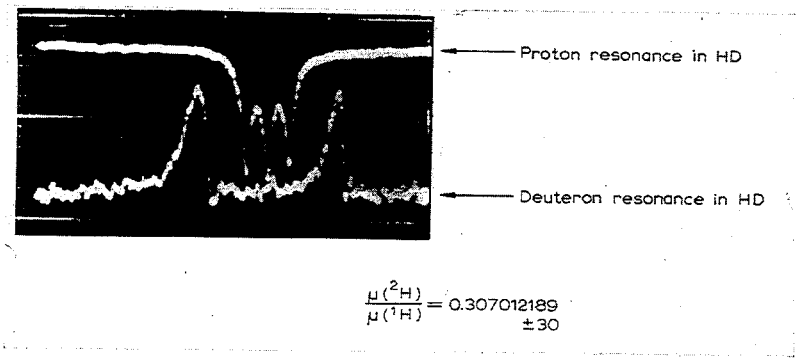


Fig. 1. Simultaneous display of deuteron and proton resonances in HD. The proton trace is inverted. (T. F. Wimett, *Massachusetts Institute of Technology.*)

and heavy water, H_2O and D_2O, might serve as our sample. Then f_D/f_H is the ratio of the magnetic moment of the deuteron to that of the proton, apart from the simple factor I_D/I_H, which is the ratio of the spins, in this case 2. In this way the magnetic moments of many different nuclei have been measured, relative to the moment of the proton, with great precision.

To show what accuracy can be achieved, let me describe one example, a very recent experiment by T. F. Wimett in the laboratory of Professor Bitter of the Massachusetts Institute of Technology. The ratio of the deuteron moment to the proton moment had been carefully determined before by Anderson and Smaller at Chicago, and by K. Siegbahn and Lindström here at the Nobel Institute. Wimett has now refined this measurement still further using as his sample the gas HD. This substance has a certain advantage: the two different nuclei are not only in the same vessel, but in the same molecule, so there is no doubt that the magnetic field at each is the same. Fig. 1 shows simultaneous traces of the proton resonance and the deuteron resonance in a field of about 15,000 gauss. You will notice at once that the resonances are not simple. The proton resonance is a triplet line, almost fully resolved, while the deuteron resonance is a cleanly resolved doublet. This has an interesting explanation, to which I shall return near the end of my lecture, but it has no direct bearing on this experiment except to emphasize the extraordinary resolution. The separation of the deuteron doublet is only 0.07 gauss, 1 part in 200,000 of the applied field. By carefully lining up the central proton peak with each deuteron peak, in turn, a measurement reliable to about 1 part in 10 million can be made. The determination of the exact frequency ratio is accomplished by an elaborate scheme of harmonic mul-

tiplication. Wimett's result, which he kindly gave to me just a few days before I left Cambridge, is

$$\frac{\mu(\mathrm{D})}{\mu(\mathrm{H})} = 0.307012189 \pm 0.000000030$$

Table 1. Magnetic moments of the neutron (n), deuteron (^{2}H), triton (^{3}H), and ^{3}He, relative to the moment of the proton (^{1}H).

$\frac{\mu(\mathrm{n})}{\mu(^1\mathrm{H})} = -0.685001 \pm 3$	$\frac{\mu(^2\mathrm{H})}{\mu(^1\mathrm{H})} = 0.30701225 \pm 10$
$\frac{\mu(^3\mathrm{H})}{\mu(^1\mathrm{H})} = 1.066636 \pm 10$	$\frac{\mu(^3\mathrm{He})}{\mu(^1\mathrm{H})} = -0.7618150 \pm 12$

The results of four earlier experiments of this sort are listed in Table 1. Of course the neutron experiment is a very special one, and Professor Bloch has already described his elegant method for detecting the neutron precession frequency. The other values were obtained by various investigators[1] in the way I have just indicated. These numbers provide a very stringent test of any theory of nuclear structure. A wholly satisfactory theory ought to be able to predict them. As you know, we are very far from having such a theory now. But even so, the high accuracy is not all wasted. The moment of the deuteron is not *quite* equal to the sum of the moments of its separate parts, the neutron and proton. Thanks to the extreme precision of the experimental values, this small but significant discrepancy is itself fixed quite accurately.

As we leave the simplest nuclei and proceed through the Periodic Table, the situation changes a little. The experimental precision is just as high, but one must apply to the measured frequency ratios a small correction to take account of the magnetic shielding effect of the atomic electrons. This correction increases rapidly with atomic number. Until better atomic wave functions are available, it cannot be calculated reliably enough to support the six figure accuracy of the resonance measurement. There is one important exception. For two isotopes of the same element, the correction cancels out. Accurate moment ratios for isotopic pairs are of considerable interest, especially if the hyperfine structure interval for the two isotopes can be determined. A. Bohr and Weisskopf[2] have shown that such a comparison can reveal something about the internal structure of the nucleus.

Perhaps the greatest present need, through most of the Periodic Table, is for spin values and only moderately precise moment values, to help test the promising shell theories of nuclear structure. Nuclear induction methods will surely continue to contribute to this task. Spin determinations in particular, which depend on careful intensity measurements, are likely to receive more attention.

The experimental physicist often needs to measure the intensity of a magnetic field. Until recently a precise magnetic field measurement has been a formidable undertaking, but it is such no longer. At the National Bureau of Standards in Washington, Hipple and his collaborators[3] have measured the proton precession frequency and, at the same time, have measured in absolute units the intensity of their magnetic field. All the resources of the Bureau were brought to bear on the latter measurement, and an accuracy of 1 part in 40,000 was achieved. Knowing their result, an experimenter anywhere in the world can determine absolute magnetic intensities to the same precision, using equipment no more elaborate than an ordinary radio receiver. He need only determine the proton precession frequency in the field in question. Few physical quantities are as easy to measure accurately as a radio frequency, thanks to modern electronic techniques and the availability, in most countries, of standard-frequency broadcasts.

Already a number of experiments have been performed in which the nuclear resonance has served as a standard of reference. Certain ones are of special interest because they have improved our knowledge of the fundamental atomic constants. In each of these experiments two *different* magnetic resonance phenomena have been observed at the same time in the same magnet. The phenomena involved are indicated in Fig. 2. The precession of the proton moment (Fig. 2 a) we have already described; it is detected and measured in a nuclear induction experiment. The electron has an intrinsic spin and magnetic moment and also precesses in a magnetic field (Fig. 2 b). For the same field strength, the electron's spin-precession frequency is about 700 times greater than the proton precession frequency. A bare proton, moving in a magnetic field, revolves in a circular orbit with a certain frequency (Fig. 2 c). This is the familiar principle of the cyclotron and we might name the associated frequency the « proton cyclotron frequency ». A free electron in a magnetic field likewise revolves in a circular orbit, with a frequency that we may call the « electron cyclotron frequency ». These « cyclotron frequencies » are governed simply by the particle's charge-to-mass ratio. If any pair of these resonance phenomena are compared, in the same magnetic field,

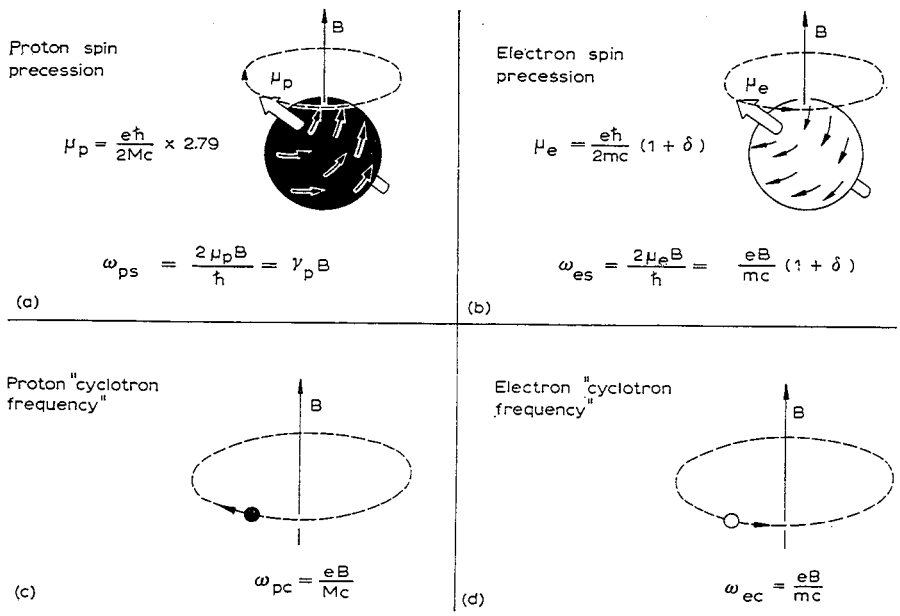


Fig. 2. Four elementary magnetic resonance phenomena.

the field strength cancels out and we are left with a relation between certain atomic constants. The experiment is thus reduced to the measurement of a frequency *ratio*.

The ratio of the proton spin-precession frequency to the proton cyclotron frequency has been determined by Hipple[4], and by Bloch[5], in ways very different but equally ingenious. (The direct result of this measurement is the value of the proton moment in nuclear magnetons.) In my laboratory we have measured the ratio of the proton precession frequency to the cyclotron frequency of free electrons[6]. The precession of the spin of a truly free electron has not been observed, but Kusch and his collaborators at Columbia, by combining the techniques of atomic beams and nuclear induction, have determined very precisely the ratio of the proton's precession frequency to that of the spinning electron in the hydrogen atom[7]. The results of these experiments have in one way or another improved the accuracy of the following constants of atomic physics: the Faraday, F; the specific charge of the electron, e/mc; the ratio of the mass of the proton to that of the electron, M/m; the dimensionless *fine-structure constant*, $2\pi e^2/hc$. They have also helped to test the new theoretical advances in quantum electrodynamics led by Schwinger.

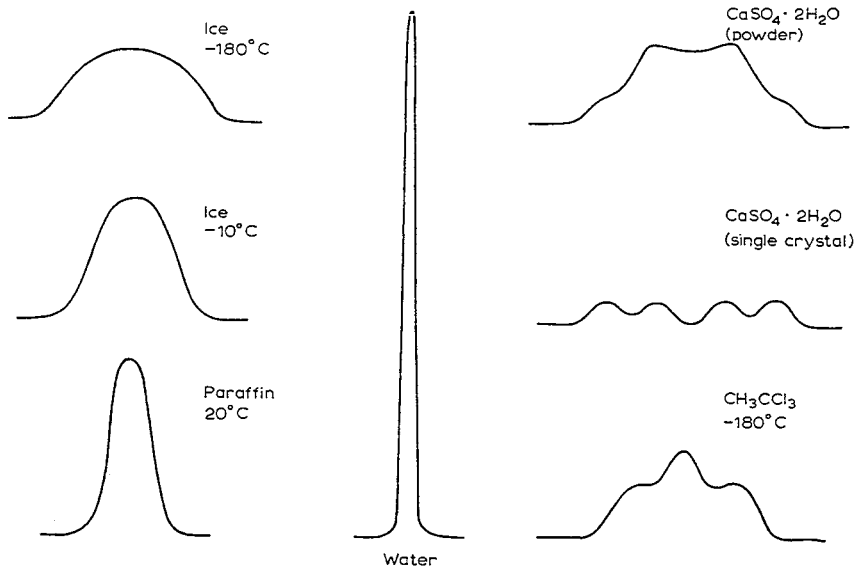


Fig. 3. Line shapes typical of the proton resonance in various substances. The line in water is not drawn to scale; it is actually much narrower and more intense than indicated.

We turn now to a very different subject, one whose rapid growth was surely not anticipated at the beginning of this work. That is the study of nuclear magnetism for the light it can throw on problems of molecular structure and molecular motion, problems rather close to physical chemistry. Indeed certain branches of this work are now being pursued in chemical laboratories. We are interested not in the mere occurrence of nuclear resonance at a particular frequency but in its finer details. The variety that we find is suggested by Fig. 3. Here are shown several line shapes typical of the proton magnetic resonance in a few simple substances. One way to record such a spectrum is to hold the applied magnetic field constant while slowly varying the frequency of observation. In every case the center of the resonance absorption occurs at very nearly the same frequency, in these substances. The line width, on the other hand, varies from some 50 kilocycles/second, in the case of ice at low temperature, to less than a few cycles/second, in the case of a liquid like water. (The true width of the resonance in water is too small to be shown properly in the diagram.)

The effects shown here can all be traced to the action of the nuclear magnets on one another. In ice, each hydrogen nucleus is subjected not only to the field of our magnet, which might be several thousand gauss, but to the

small magnetic field caused by each nearby proton in the crystal. Each nearest neighbor produces a field of a few gauss, more or less random in direction, with the result that the resonance absorption for the crystal as a whole is spread out over a corresponding interval in frequency. These « local magnetic fields » are present in water, too, but their effectiveness is destroyed by the rapid molecular diffusion characteristic of a liquid. Two adjoining molecules remain neighbors so short a time (about 10^{-11} seconds) that their mutual influence is very slight. Thus *molecular motion* is the key to the striking difference between crystals and fluids in respect to the width of the nuclear magnetic resonance.

In many crystals, atoms or molecules occasionally jump from one position to another. Although such events are rare by most standards, they may occur frequently enough to affect the width of the nuclear resonance line. In ice, for instance, the line is distinctly narrower at $-10°C$ than at a much lower temperature. The narrowing indicates that each hydrogen atom undergoes a displacement within the crystal lattice several thousand times per second. The motion involved here is not the familiar thermal vibration of an atom in a perfectly elastic crystal, but an abrupt shift to a new site. Nuclear resonance studies have disclosed many unsuspected internal motions in crystals, and have helped to reveal the exact nature of motions already suggested by other physical or chemical evidence[8].

If the magnetic nuclei in a crystal are clustered together in pairs, the action of the nuclear magnets on each other causes a doubling of the magnetic resonance line. Further distortion is caused by more remote neighbors, and the width of the pattern as a whole depends on the orientation of the nuclear pair with respect to the magnetic field. Rather complicated effects are possible, some of which are illustrated in Fig. 3. From a complicated effect that is also well understood, one can usually get a good deal of information. These magnetic interactions are now rather well understood. I shall mention only one recent application of the technique to a chemical problem.

Groups of three equidistant protons in randomly oriented crystals give, characteristically, a line with three humps. The methyl chloroform line, in Fig. 3, is an example. We have a complete quantitative theory of this line structure[9]. Last year Richards and Smith[10] at Oxford, and quite independently Kakiuchi and others[11] in Tokyo, studied the proton resonance in perchloric acid monohydrate ($HClO_4 \cdot H_2O$) and certain related substances. This type of line was found. They were able to prove thereby that in those crystals the « hydronium » ion, OH^+, exists, and to determine its dimensions.

It is well known that hydrogen atoms are very hard to locate by X-ray diffraction. Here is an opportunity for nuclear resonance studies to supplement, in special cases, the more general methods of structure analysis.

Many nuclei have, in addition to an intrinsic magnetic dipole moment, an *electric quadrupole moment*. This is a way of saying that the electric charge of the nucleus is not distributed evenly over a sphere. One may think of the nucleus as either an elongated or a flattened ellipsoid. If such a nucleus is put into a non-uniform electric field, it experiences a torque, and precesses. (You will recall that the slow precession of the earth's axis is a consequence of the ellipticity of the earth and the non-uniform gravitational fields of the moon and sun.) The nuclei in many crystals find themselves in non-uniform electric fields arising from their immediate atomic surroundings. A nuclear electric quadrupole moment then manifests itself in a splitting of the magnetic resonance line. To put it another way, the energy associated with each of the possible orientations of the nuclear spin in a magnetic field is somewhat altered by the electric effect. This is illustrated in Fig. 4. The trace in Fig. 4 is the nuclear resonance spectrum of ^{23}Na in a crystal of sodium nitrate. The line is widely split into a triplet by the quadrupole interaction. The number of components is always twice the nuclear spin I, so we have here a way of determining nuclear spins. This example is due to R. V. Pound who was the first to investigate these effects in crystals[12].

Another interesting and fruitful branch of nuclear magnetism is the study of nuclear relaxation, by which I shall mean here the attainment of thermal equilibrium between the nuclear spins and their environment. Professor Bloch has already referred to the crucial importance of relaxation in all these experiments. In the first experiment of Torrey, Pound, and myself this question gave us much concern. Our approach was somewhat different from Bloch's and was based on a theoretical paper of fundamental importance by Professor Waller[13]. Waller, in 1932, was dealing with relaxation in electronic paramagnetism, but Torrey was able to adapt his results to our case, with the conclusion that equilibrium should certainly be reached within several hours. We then designed the experiment so that, even if so long a time were required, the resonance could still be detected. As it turned out, the precaution was unnecessary, but that was our introduction to a fascinating problem that has occupied our attention ever since.

I believe we now have a good understanding of nuclear relaxation in many – though not all – physical systems. Experimentally, we find that the time to establish thermal equilibrium ranges widely from hours, in some substances,

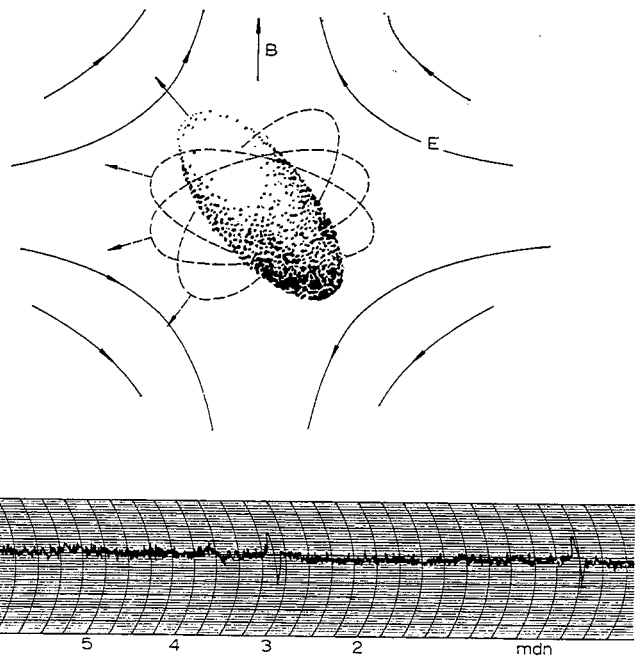


Fig. 4. A nucleus with an electric quadrupole moment in a uniform magnetic field (B) and a non-uniform electric field (E). *Lower trace*: ^{23}Na resonance in a single crystal of $NaNO_3$, showing splitting of magnetic resonance line into a triplet.

down to fractions of a millisecond in others. I show only one example, and that mainly to clarify the meaning of the term relaxation. Fig. 5 is a plot of the gradual approach to equilibrium magnetization of an originally unmagnetized crystal of ammonium bromide. This can be called, quite properly, a *cooling curve*. The nuclear spins, in order to align themselves with the magnetic field, must give off energy to the crystal lattice. They are coupled so weakly to the lattice that the transfer of energy takes more than several minutes. Actually the approach to equilibrium is exponential, and the corresponding time constant, in this example 32 seconds, is what one calls the spin-lattice relaxation time.

It may seem astonishing, at first, that so long a time is associated with an atomic process. But in fact, if one looks at the problem more closely, it is hard to understand why the time is not much *longer*. Moreover this crystal, at a somewhat higher temperature, has a relaxation time of only 0.01 seconds! The explanation of this, and of similar behavior in a wide class of substances, lies in internal molecular motions other than mere vibration. It is

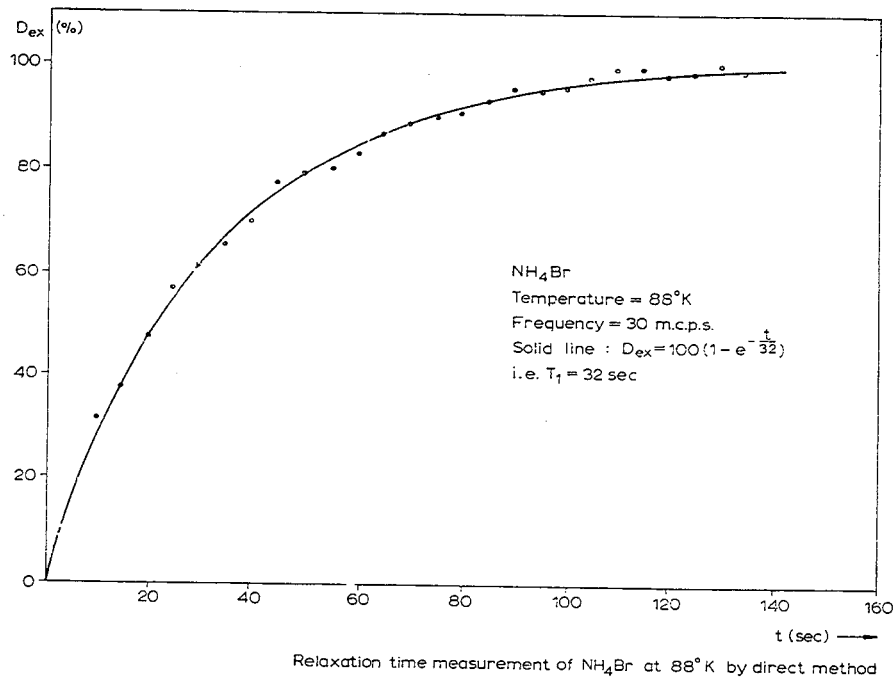


Fig. 5. The gradual approach to equilibrium magnetization. The ordinate is proportional to the intensity of nuclear polarization.

closely connected with the cause of line-narrowing, mentioned earlier. The particular motion occurring in the ammonium halides is a sudden rotation of the whole tetrahedral ammonium ion inside the cubic unit cell.

We have now a fairly general theory of nuclear relaxation[14], a theory that has proved reliable enough to allow one to draw conclusions about the nature of the internal motion in a substance from the observed nuclear relaxation. My early collaborator, N. Bloembergen, has made essential contributions to this subject. Experimental work at Oxford and at Leiden has added much to our knowledge of other aspects of nuclear relaxation. In many respects this is only a new branch of a somewhat older subject, paramagnetic relaxation, in the study of which the Dutch laboratories, and especially Professor Gorter, have long led. The problem of nuclear relaxation has a special attraction for the low-temperature physicist, because, as Professor Simon pointed out many years ago, nuclear magnetization offers a way toward extremely low temperatures, providing the nuclear spins can exchange heat with something else.

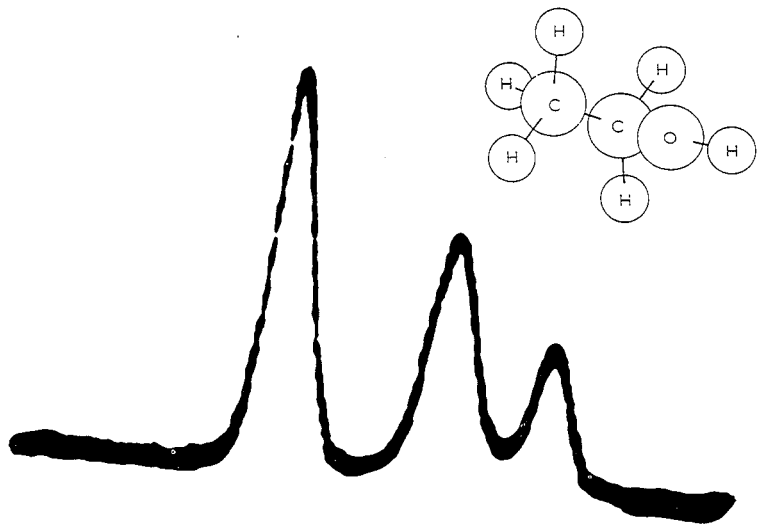


Fig. 6. The proton resonance in ethyl alcohol, observed with high resolution. The three lines arise from the CH_3 hydrogens, from the CH_2 hydrogens, and from the OH hydrogen, respectively.

There remain several puzzling aspects of nuclear relaxation. In trying to understand them we find ourselves still returning to some of the ideas advanced twenty years ago by Professor Waller. While I speak of theoretical contributions to nuclear magnetism, I must mention also Professor Van Vleck, who has put the theory of line-width on a rigorous basis, a notable advance[15].

It is an old story in physics that higher resolving power leads to new effects. We remember that the magnetic moment of the nucleus was itself discovered through the hyperfine structure of lines in the visible spectrum. The nuclear resonance line in a liquid or gas can be remarkably narrow, as you have already seen. As soon as the reason for this was recognized, it became clear that the only practical limit on resolution was the inhomogeneity of the magnetic field applied to the specimen. Efforts were made in many laboratories to improve the magnets, and to use smaller specimens as well. With the improved resolution, it was found that identical nuclei, in the same applied field but in chemically different molecules, do not precess at exactly the same frequency. The explanation is simple: the magnetic field at an atomic nucleus differs slightly from the field externally applied because of the shielding effect of the electron cloud around the nucleus. In different molecules the atom's electron configuration will differ slightly, reflecting differences in

the chemical bond. The resulting resonance shifts have been called « chemical shifts ». They are only a nuisance to the experimenter interested in exact ratios of magnetic moments. But they are interesting to the physical chemist because they reveal something about the electrons that partake in the chemical bond.

No laboratory has more assiduously pursued high resolution than Professor Bloch's, where the striking example shown in Fig. 6 was discovered[16]. This is the proton resonance in ethyl alcohol, seen under very high resolution. Each lines comes from a chemically equivalent group of protons in the molecule, the intensity being proportional to the number of hydrogen atoms involved, 3 (methyl group), 2 (CH_2 group), and 1 (hydroxyl), respectively.

I can now return to the splitting observed in the proton and deuteron resonances in Fig. 1. This splitting does not arise from any effect on the applied field, but from an indirect coupling of the proton to the deuteron through the medium of the two electrons in the HD molecule. It is as much an intrinsic property of the HD molecule as is the optical spectrum of the molecule[17]. The splitting amounts to 43 cycles per second. I am sure we have only begun to explore the domain of very weak interactions – the « audio spectrum » of molecules, if I may call it that.

This has been a long story and a complicated one. We are dealing not merely with a new tool but with a new subject, a subject I have called simply nuclear magnetism. If you will think of the history of ordinary magnetism – the electronic kind – you will remember that it has been rich in difficult and provocative problems, and full of surprises. Nuclear magnetism, so far as we have gone, is like that too.

1. B. Smaller, *Phys. Rev.*, 83 (1951) 812; F. Bloch, A. C. Graves, M. Packard, and R. W. Spence, *Phys. Rev.*, 71 (1947) 551; H. L. Anderson, *Phys. Rev.*, 76 (1949) 1460.
2. A. Bohr and V. Weisskopf, *Phys. Rev.*, 77 (1950) 94.
3. H. A. Thomas, R. L. Driscoll, and J. A. Hipple, *Phys. Rev.*, 78 (1950) 787.
4. H. Sommer, H. A. Thomas, and J. A. Hipple, *Phys. Rev.*, 80 (1950) 487.
5. F. Bloch and C. D. Jeffries, *Phys. Rev.*, 80 (1950) 305.
6. J. H. Gardner and E. M. Purcell, *Phys. Rev.*, 76 (1949) 1263.
7. S. H. Koenig, A. G. Prodell, and P. Kusch, *Phys. Rev.*, 88 (1952) 191.
8. H. S. Gutowsky, G. B. Kistiakowsky, G. E. Pake, and E. M. Purcell, *J. Chem. Phys.*, 17 (1949) 972.
9. E. R. Andrew and R. Bersohn, *J. Chem. Phys.*, 18 (1950) 159.

10. R. E. Richards and J. A. S. Smith, *Trans. Faraday Soc.*, 47 (1951) 1261.
11. Y. Kakiuchi, H. Shono, H. Komatsu, and K. Kigoshi, *J. Phys. Soc. Japan*, 7 (1952) 102.
12. R. V. Pound, *Phys. Rev.*, 79 (1950) 685.
13. I. Waller, *Z. Physik*, 79 (1932) 370.
14. N. Bloembergen, E. M. Purcell, and R. V. Pound, *Phys. Rev.*, 73 (1948) 679.
15. J. H. van Vleck, *Phys. Rev.*, 74 (1948) 1168.
16. J. T. Arnold, S. S. Dharmatti, and M. E. Packard, *J. Chem. Phys.*, 19 (1951) 507.
17. N. F. Ramsey and E. M. Purcell, *Phys. Rev.*, 85 (1952) 143.

Biography

Edward Mills Purcell was born in Taylorville, Illinois, U.S.A., on August 30, 1912. His parents, Edward A. Purcell and Mary Elizabeth Mills, were both natives of Illinois. He was educated in the public schools in Taylorville and in Mattoon, Illinois, and in 1929 entered Purdue University in Indiana. He graduated from Purdue in electrical engineering in 1933.

His interest had already turned to physics, and through the kindness of Professor K. Lark-Horovitz he was enabled, while an undergraduate, to take part in experimental research in electron diffraction. As an Exchange Student of the Institute of International Education, he spent one year at the Technische Hochschule, Karlsruhe, Germany, where he studied under Professor W. Weizel. He returned to the United States in 1934 to enter Harvard University, where he received the Ph.D. degree in 1938. After serving two years as instructor in physics at Harvard, he joined the Radiation Laboratory, Massachusetts Institute of Technology, which was organized in 1940 for military research and development of microwave radar. He became Head of the Fundamental Developments Group in the Radiation Laboratory, which was concerned with the exploration of new frequency bands and the development of new microwave techniques. This experience turned out to be very valuable. Perhaps equally influential in his subsequent scientific work was the association at this time with a number of physicists, among them I. I. Rabi, with a continuing interest in the study of molecular and nuclear properties by radio methods.

The discovery of nuclear magnetic resonance absorption was made just after the end of the War, and at about that time Purcell returned to Harvard as Associate Professor of Physics. He became Professor of Physics in 1949; his present title is Gerhard Gade University Professor. He has continued to work in the field of nuclear magnetism, with particular interest in relaxation phenomena, related problems of molecular structure, measurement of atomic constants, and nuclear magnetic behaviour at low temperatures. He has made some contribution to the subject of radioastronomy.

He is a Fellow of the American Physical Society, a member of the National

Academy of Sciences, of the American Academy of Arts and Sciences, and of the President's Science Advisory Committee under President Eisenhower from 1957–1960 and under President Kennedy as from 1960.

In 1937, Purcell married Beth C. Busser. They have two sons, Dennis and Frank.

Physics 1953

FRITS ZERNIKE

«for his demonstration of the phase-contrast method, especially for his invention of the phase-contrast microscope»

Physics 1953

Presentation Speech by Professor E. Hulthén, member of the Nobel Committee for Physics

Your Majesty, Your Royal Highnesses, Ladies and Gentlemen.

The Royal Academy of Sciences has this year awarded the Nobel Prize for Physics to Professor Frits Zernike, Groningen, for the phase-contrast method devised by him, and particularly for his invention of the phase-contrast microscope.

Zernike's discovery falls within that part of optics in which one operates with the notion of light as a wave motion. From this it follows, amongst other things, that light may be extinguished by light through interference, and is diffracted and scattered by small particles such as the microscopic objects. All this, as far as the principles involved are concerned, belongs to a closed chapter that is generally referred to as classical physics.

When on this occasion a Nobel Prize is awarded for contributions in classical physics, the fact is so remarkable that we must go back to the very earliest Nobel Prizes to find a counterpart. All later Nobel Prizes, with the exception of a couple of awards where the stress was rather upon the technical aspect, have been awarded for discoveries in atomic and nuclear physics, the physics of this century.

It would scarcely be an exaggeration to claim that the microscope is one of our most important instruments of research. Every improvement, even a slight sharpening of this eye towards the microcosmos, may pave the way to great advances in the natural sciences, medicine, and the technical sciences.

Probably no other instrument has been the object of so much technical and theoretical study as the microscope. The thorough theoretical foundation that we owe to the genius of Ernst Abbe of the famous Zeiss concern was followed at the end of the last century by a development of the microscope that brought its optical and illumination system very close to perfection.

But even Abbe's theory had a gap, for it took into account only those conditions in which the microscopic objects appear against the background as a result of their contrasts in colour and intensity. Many microscopic ob-

jects, however, micro-organisms such as bacteria and cells, are colourless and transparent, and for this reason difficult to distinguish from their surroundings. Attempts have been made to overcome this difficulty with various methods of staining or with a special illumination system, the so-called dark-field illumination. The staining methods are not always suitable, as for example when we are dealing with living objects; and dark-field illumination easily leads to a misinterpretation of the finer details in the structural picture.

It was this gap in Abbe's theory that in the 1930's led Zernike to re-investigate the refraction processes in the light that give rise to the image in a microscope. Even if the eye is not able to discern the change undergone by a beam of light when it passes through a transparent object, the change does nonetheless exist as a phase-difference of a quarter of a wavelength relative to the direct beam that does not pass through the object. The problem was thus to transform these otherwise imperceptible phase differences to visible contrasts in intensity. Zernike was able to show that this was possible, thanks to the fact that the two rays of light take different routes through the microscope before being reunited in the image. By interposing in the paths of the direct ray a so-called « phase-plate », which either further increases the phase-displacement to half a light-wavelength or smooths it out completely, Zernike attained the desired effect, so that the two rays either extinguish or reinforce each other. In this way the formerly invisible particle appears in dark or light contrast to the surroundings.

I have deliberately dwelt upon the description of the phase-contrast microscope as the result of Zernike's method which is, so far, the most valuable. The phase-contrast method has, however, many other and increasingly important applications in optics. In addition to its capacity to render colourless and transparent objects visible in the microscope, it also enables one to detect slight flaws in mirrors, telescope lenses, and other instruments indispensable for research. In this connection, Zernike's phase-plate serves as an indicator which locates and measures small surface irregularities to a fraction of a light-wavelength. This sharpness of depth is so great that it penetrates to the point at which the atomic structure of the substance begins to become manifest.

Professor Zernike. The Royal Academy of Sciences has awarded you the Nobel Prize in Physics for your eminent « method of phase contrast » and especially for your invention of the « phase-contrast microscope ».

I now ask you to receive the prize from the hands of His Majesty.

FRITS ZERNIKE

How I discovered phase contrast

Nobel Lecture, December 11, 1953

« Phase contrast » was not discovered while working with a microscope, but in a different part of optics. It started from my interest in diffraction gratings, about from 1920 on. Such a (reflecting) grating consists of a plane or concave mirror with a large number of equidistant grooves ruled on its surface. Small nearly unavoidable imperfections in the location of the grooves show clearly in the optical behaviour of the grating. The most conspicuous error is a periodic one which repeats itself after each revolution of the screw of the ruling engine. The regularly recurring displacement of the grooves causes corresponding changes of the optical path, just as if the mirror surface were wavy. Consequently the instrument behaves as if a coarse grating, with a constant of about 2 millimeters, were superimposed on it, so that each strong spectral line is accompanied to right and left by a number of weak spurious lines, the so-called « Rowland ghosts ». These have a remarkable effect if one looks down at the surface of the grating, placing the eye at the position of a spectral line. A perfect grating would in this case show an evenly illuminated surface, in the colour of the spectral line. In reality, however, one sees a strongly striped surface. At the end of a 1902 paper, H. S. Allen remarked that these stripes were nothing real, but simply the effect of the interference between the principal line and its ghosts. Indeed the stripes disappear when the ghosts are covered up. I remember strongly objecting against his conclusion of unreality. On the contrary I was convinced that the striped surface gave more information about the periodic ruling errors than that obtainable by photographing the ghosts, because in the first case the relative phases of the gosts come into play, whereas these are lost in the second case. I kept the question in mind, planning to look further into it as soon as an opportunity would arrive.

About 1930 our laboratory had obtained a large concave grating and set it up in a Runge-Paschen mounting. The striped appearance of the surface was soon found, but as the grating was 6 meters from the eye, I tried a small telescope pointed at the grating. Then the unexpected happened. The stripes were seen very clearly, but disappeared as the telescope was exactly focussed

on the surface of the grating! By a succession of experiments and calculations I soon succeeded in explaining this. On looking back to this event, I am impressed by the great limitations of the human mind. How quick are we to learn, that is, to imitate what others have done or thought before. And how slow to understand, that is, to see the deeper connections. Slowest of all, however, are we in inventing new connections or even in applying old ideas in a new field. In my case the really new point was that the ghosts differed in phase from the principal line. Now it is common knowledge that in all interference phenomena differences of phase are all-important. Why then had phases never been considered before in this case, nor in the corresponding one in the microscope? Some excuse may be found in the difficulty to define them exactly. Let me explain this for a more simple case, the diffraction image of a slit. The way to observe this may be as follows. A telescope is pointed at a vertical line-source of light, such as the filament of an incandescent lamp. A vertical slit of say 2 mm width is placed closely behind the objective of the telescope. This causes the image of the source to be broadened out into a diffraction pattern: a bright central stripe (order zero) is accompanied on both sides by weaker and weaker secondary maxima (orders one, two, etc.). The formula for this diffraction pattern is given in the textbooks, the amplitude being determined by the function $\sin x/x$. In the few cases where the phases are mentioned in the literature, on the other hand, there is no consensus of opinion. Some say the phases are equal over the whole pattern – except for the obvious reversal of the odd orders, whereas others make them change proportional to x^2. I find that it all depends on the, often tacitly assumed, surface to which the phases are referred. If this reference surface is the focal plane of the telescope objective, one comes to the second statement, if it is a cylindrical surface with the centre line of the slit as axis, the equality of phases results.

You may want to ask whether these phases can be observed. I find they can. All one has to do is to throw the diffraction image on a *coherent background*, obtained in the following way. The slit is covered by a glass plate with a thin metallic layer which transmits a few percent of the light. A fine scratch is made in this layer, forming a narrow slit which is adjusted so as to lie in the centre of the broad slit. The light through the scratch is broadened out by diffraction and thus forms the desired background, which interferes with the diffraction pattern. The phases of this pattern are thus compared with those of the auxiliary wave forming the background. In the experiment the auxiliary wavefront therefore plays

the role of the cylindrical reference surface in the theoretical treatment.

It is only by the introduction of an adequate reference surface that a definite statement about the phase differences involved can be made. In the case of the Rowland ghosts the result was: their phases differ by ninety degrees from the principal line. Now I happened to know of a simple method to change this. Lord Rayleigh described in 1900 how to make very shallow etchings in glass surfaces without spoiling their optical quality, by the slow action of very dilute hydrofluoric acid. By this process I made what I called *phase-strips:* glass plates with a straight groove, a millimeter or less wide and of a uniform depth of half a wavelength. Such a phase-plate was placed in the spectrum so that a bright spectral line fell on the strip, whereas its ghosts passed through the glass beside it. In a telescope behind the phase-plate the stripes on the grating surface then stood out clearly.

For a physicist interested in optics it was not a great step to change over from this subject to the microscope. Remember that in Abbe's remarkable theory of the microscope image the transparent object under the microscope is compared with a grating. To be precise, a transmission grating is considered as the test-object and the diffraction by this grating as the primary phenomenon. At first sight this has nothing to do with the magnified image of the object formed by the microscope objective. Instead, the objective forms an image of the light source, practically in its back focal plane, consisting of a central direct image accompanied by diffracted images on both sides. This, although on a very much smaller scale, is the analogue of the grating line with its ghosts. The light issuing from these images overlaps in the eyepiece of the microscope and by interference gives rise to stripes which, curiously enough, resemble a magnified image of the object! Abbe's theory has been summarized in this sentence: «The microscope image is the interference effect of a diffraction phenomenon.»

You will now readily understand that, acquainted with this theory, I soon tried my phase-strip in a microscope, throwing the direct image of a linear light-source on the strip placed closely above a low-power objective.

I must now explain why the unexpected discovery of the 90° phase shift applies to the microscope image as well. It all depends on the nature of the object under the microscope. In his theory Abbe and his followers always considered an object of alternate opaque and transparent strips. The diffraction images for such a grating calculated in the well-known way are in phase with the central image. On the other hand, if the object consists of alternate thicker and thinner strips, then the phase difference of 90° is found. In the

first case, the diffraction is caused by the unequal *amplitudes* of the light passing the strips, in the second case by the unequal light paths, i.e. by the unequal *phases*. I therefore distinguish the two by calling the first kind an *amplitude grating*, the second a *phase grating*. Or in the general case of an irregular structure, an « amplitude object », resp. a « phase object ». Nearly all objects of biological or medical interest belong naturally in the second group. The highly developed staining techniques evidently aim at changing them, or their special details one wants to see, into amplitude objects.

It will now be seen that for a phase object my phase-strip in the focal plane of the microscope objective brought the direct image of the light-source in phase with the diffracted images, making the whole comparable to the images caused by an amplitude object. Therefore the image in the eyepiece appears as that of an absorbing object, that is with black-and-white contrast as if the object had been stained. The full name of the new method of microscopy might be something like: « phase-strip method for observing phase objects in good contrast ». I shortened this into *phase-contrast method*. Before going into further details about the development of the method, a few general remarks should be made.

In a treatise on the Abbe theory, Lummer comes to the conclusion that « in the ideal case the microscope image is exactly similar to the object in structure and phase ». Now the absolutely transparent details of a phase object leave the intensity of the passing light unchanged. All they do is to impress phase differences on it. According to Lummer, then, the image will show the same phase differences – which however are invisible – and an equal intensity everywhere. In other words, the phase object is absolutely invisible « in the ideal case ». Of course the practical microscopist has never been content with this; as a matter of fact, he never found it out! Without realizing it, he had always turned the fine adjustment, that is, put the object a little out of focus, in order to see the tricky transparent details. Only a somewhat diffuse and watery image is obtained in this way. This will be explained further on.

With the phase-contrast method still in the first somewhat primitive stage, I went in 1932 to the Zeiss Works in Jena to demonstrate. It was not received with such enthusiasm as I had expected. Worst of all was one of the oldest scientific associates, who said: « If this had any practical value, we would ourselves have invented it long ago ». Long ago, indeed! The great achievements of the firm in practical and theoretical microscopy were all due to their famous leader Ernst Abbe and dated from before 1890, the year in which Abbe became sole proprietor of the Zeiss Works. From then on he

had been absorbed in administrative and social problems, partly also in other fields of optics. Indeed his last work on microscopy dates from that same year. In it he gave a simple reason for certain difficulties with transparent objects, but this was of no account. His increasing staff of scientific collaborators evidently under the impression of his inspiring personality, formed the tradition that everything worth knowing or trying in microscopy had been achieved already. For more than twenty-five years after Abbe's death in 1906, his great authority thus barred the way to further progress.

Here is one more remarkable historic point. Whereas all the other achievements of Abbe's were greatly appreciated by all practical microscope users, his theory of image formation was firmly rejected by most of them. To the physicist this may seem incredible, especially when he remembers Abbe's experiments which in his opinion confirm the theory in a convincing way. The opposing microscopists, however, said these experiments only showed how the microscope may be used, or rather misused, by the physicist for interference experiments which have nothing to do with the ordinary proper use of the instrument. A long story could be told about the violent controversies of this kind which occurred time and again through half a century. We can say now that Abbe and his followers were as much to blame as the other party. For one thing, their theory was too abstract, it only applied to the oversimplified cases of a point source of light and an object of regular structure. However, it could very well have been extended to meet the practical cases. But then it was also incomplete, it did not explain the peculiarities in the imaging of transparent objects, and what is worse, its defenders never recognized this incompleteness. Small wonder therefore that the microscopist rejected the theory as useless in practice.

Returning now to the phase-contrast method, I will first give an account of its working principle. Let the incident light for simplicity be a plane wave. Without an object, that is if there is only a clear glass-plate under the microscope, this wave passes unchanged, is brought to a focus closely above the objective (in its back focal plane), and spreads out again to an evenly illuminated field in the eyepiece. If there is a real object, every small detail of it will give rise to a slight perturbation of the wave. One may always consider this as resulting from a perturbed wave to be superimposed – in amplitude, not in energy – on the unchanged wave. This last one shall be called the *direct light*, it will clearly give the even background. The perturbed wave will spread out from the detail in all directions, will fill the whole aperture of the objective and will reunite in the corresponding image point in the eyepiece.

The perturbed waves from all the object points together will be called the *diffracted light*. The microscope image in the eyepiece now results from the interference of the diffracted light with the direct light. In order to obtain phase contrast the two must be treated differently, so as to change their relative phases. This is possible, because they are spatially separated in the back focal plane of the objective. The interplay of phases in this decomposing and reuniting of vibrations can best be visualized in a vector diagram (Fig. 1a). As is well known, a harmonic vibration is obtained from a vector MV rotating uniformly round M. The projection P on the horizontal axis performs the vibration. The vector MP′, projection of MV′ which remains always perpendicular to MV, performs a similar vibration, one quarter period in advance of P. In accordance with general usage the projecting is understood, we speak of the vibrations MV, MV′, etc.

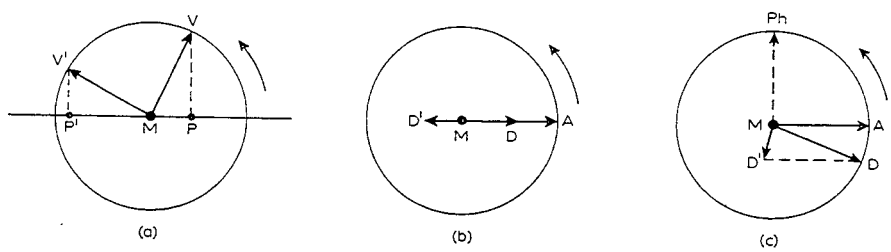


Fig. 1.

Now consider a microscopic object with slightly absorbing details on a transparent background (stained preparation). The vibration emerging from the background may be represented by MA (Fig. 1b). An absorbing detail weakens the light, it gets a smaller amplitude, such as MD. The vector MD results also from compounding MA with MD′, so that MD′ represents the change caused by the details, i.e. the perturbed vibration. Now according to a well-known theorem the optical paths along all rays from an object point to its image are equal. Therefore the direct and the diffracted vibrations arrive at the image point in the same relative phases they had in the object, and the same Fig. 1b may thus serve for the reuniting of these vibrations. As a result the absorbing detail is seen darker than the background. Now compare this with the case of a transparent object (unstained preparation, Fig. 1c). Its details will ordinarily be somewhat stronger refracting than the embedding medium. This means that the light is propagated with less speed so that the emerging vibration MD will be retarded in phase compared to MA,

but equal in amplitude. The change caused by the detail is now represented by MD′, nearly perpendicular to MA. The compounding of these in the image gives again MD, equal in intensity to the background MA, the detail remains invisible. It will appear, however, on slightly defocussing, as the light-paths are no longer equal in that case, resulting in some change of respective phases. At the same time the image becomes blurred, so that the observer has to find a compromise between its disappearance from the first cause, exact focus, and from the other, fading out by lack of focus. In the phase-contrast method however, the direct light has to pass the phase-strip, which is thinner than its surround through which the diffracted light passes. The direct light is thus advanced by 90°, being then represented by MPh. This causes the detail to be represented by the vector-sum of MPh and MD′, making it darker than the background. Clearly the relations are about the same as in Fig. 1b, the transparent detail may be said to be « optically stained ».

Two further improvements of phase contrast, which I made in the first years can now be explained. One is the absorbing phase-strip. Details in the object that are very thin will cause only very small phase differences, and in Fig. 1c this corresponds to a very short vector MD′, to be compounded with MPh. The thin detail therefore appears only very little darker than its surround, i.e. with very little contrast. Now there is no simple way of increasing the amplitude MD′ of the diffracted light, but the same result, increased contrast, may be attained by diminishing the amplitude of the direct light MPh. To this end the phase-strip must not only accelerate the direct light but also partly absorb it. This is for instance obtained by a thin metallic deposit on the strip. An absorption of 75% is often used; the strip then transmits 25% of the energy, or one half of the amplitude of the direct light. The contrast is thus doubled, a quite marked effect. In my own experiments I could go down to 4% transmission, i.e. a five times enhanced contrast, the limit being set by the unavoidable stray light. It is only under specially favourable circumstances that a higher increase has been attained by the French astronomer Lyot. In his study of the minute ripples of polished lens surfaces he had independently rediscovered phase contrast and could use strips that diminished the amplitude to one thirtieth, so that ripples only one thousandth of a wavelength high showed in good contrast.

A last point to explain is the halo that is always observed surrounding objects which show great contrast. This must be ascribed to the action of the phase-strip on the diffracted light. As we saw before, the phase-strip is meant to act on the direct light only. However, the diffracted light, which

fills the whole aperture of the objective, will for a small part be intercepted by the phase-strip and this part remains inactive. To find the effect of this missing part, we consider the reverse case, that it would be the only active part. Because of the narrow strip, it would form an image of much less resolving power, i.e. blurred by diffraction. As this part is missing, the « strip image » must be subtracted, in amplitude, from the full image formed by the whole aperture. The interference with the direct light then results in a very diffuse and weak negative image, appearing as a bright halo round dark details, as a dark halo round bright details.

With the straight phase-strips used in the beginning, the halo may be disturbing because the strip image of a small detail is by diffraction only spread out in one direction, namely perpendicular to the strip. This makes small bright spots in the image appear as if marked by short crossing pencil streaks. To remedy this I soon introduced *annular strips*, which make the halo spread in all direction, so that it is much fainter and indeed quite harmless.

Zeiss in Jena, who had started with so little enthusiasm, slowly continued with the method. After some more of my visits, after some years of developing too complicated instruments and after further delay by the War, they brought out phase-contrast objectives and accessories in 1941.

I will end by showing some slides of photomicrographs, some I made with my own phase-strips, others from the Zeiss prospectus.

Biography

Frits Zernike was born in Amsterdam, 16th July 1888, as the second son in a family of six children. His father, Carl Frederick August Zernike, was teacher in mathematics and head of a primary school in Amsterdam, and was a highly gifted man having interests in many branches of science; he compiled numerous elementary books in a series of subjects, and had also articles on pedagogy to his credit. His mother, Antje Dieperink, was also a teacher of mathematics. One of his brothers also became a professor of physics, one of his sisters, married to the well-known painter Jan Mankes, was the first woman ordained in the Dutch Protestant Church, another sister is one of The Netherlands' foremost literary figures.

Frits inherited his passion for physics from his father; as a boy he already possessed an arsenal of pots, crucibles, tubes, which he scraped together with his own pocket money, or received as gifts from understanding manufacturers. At the secondary school he excelled in the scientific subjects,' and neglected topics such as history and languages, including Greek and Latin, for which later on he was obliged to pass a State matriculation test in order to be fully admitted to the University.

During these school years he devoted all his spare time to his endless experiments, entering also the realms of colour photography. His limited financial means forced him to synthesize his own ether which he required for his photographic experiments. Other results of his ingenuity were a photographic camera and a miniature astronomical observatory equipped with the clockwork of an old record player, which enabled him to take pictures of a comet. Together with his father and mother he also indulged in solving arduous mathematical problems.

He entered the University of Amsterdam in 1905, studying chemistry, with physics and mathematics as minor subjects. His early interest in mathematics appears from a prize essay on probabilities for which he obtained a gold medal of the University of Groningen in 1908. A more elaborate work on critical opalescence was similarly rewarded in 1912 by the Dutch Society of Sciences at Haarlem, which had as jury distinguished scientists of those

days: Lorentz, Van der Waals, and Haga. When asked to choose between a gold medal and an amount of money, he wrote back that he preferred the money, since he had already enjoyed the privilege of receiving a gold medal. The prize essay later formed the basis of his doctor's thesis (1915). In its theoretical part he applied Gibbs' statistical mechanics and this formed the starting-point of years of fruitful collaboration with L. S. Ornstein, who worked in the same field.

In 1913 Kapteyn, the famous Professor of Astronomy at Groningen University, invited him to be his assistant. In 1915 he got his first university teaching post, not in chemistry, not in astronomy, but as successor of Ornstein as lecturer in mathematical physics at Groningen, where he was made a full professor in 1920. His papers on statistics include a paper with J. A. Prins, introducing the g-function for the correlation of the position of two molecules in a liquid, an extensive article in the Geiger and Scheel handbook, and an approximation method in the order-disorder problem (1940). Of his experimental work, the sensitive galvanometer, manufactured since 1923 by Kipp and Sons, Delft, is well known. From 1930 on he turned to optics, developed phase contrast, wrote on imaging errors of the concave grating and on partial coherence. With the collaboration of his pupils he solved the problem of the influence of lens aberrations on the diffraction pattern at a focus (1938–1948).

It is interesting to know that his great discovery of the phase-contrast phenomenon, which he discovered one evening in 1930 in his totally black-painted optical laboratory, did not immediately receive the attention it deserved. The world-famous Zeiss factories at Jena completely underestimated the value of his phase-contrast microscope. It was not until the German *Wehrmacht* took stock of all inventions which might serve in the war that at last (in 1941) the first phase-contrast microscopes were manufactured. The grotesque situation thus arose that the German war machinery helped to develop on an industrial scale Professor Zernike's long-neglected invention while its inventor, like his fellow-countrymen, suffered under the oppression by the same German powers during the occupation of the Netherlands. After the war, other firms also took up the production of many thousands of phase-contrast microscopes, thereby providing the service to science, and in particular to medicine, which should have been effectuated some twenty years earlier.

Zernike's achievements were recognized by the Royal Microscopical Society; he was also awarded the Rumford Medal of the Royal Society

(London) and an honorary doctorate in Medicine from the University of Amsterdam.

Zernike married twice. His first wife, Dora van Bommel van Vloten, died in 1944; they had one son. In 1954 he married Mrs. L. Koperberg-Baanders. After his retirement from Groningen University they moved to Naarden, a town in the countryside near Amsterdam.

Physics 1954

MAX BORN

«for his fundamental research in quantum mechanics, especially for his statistical interpretation of the wave-function»

WALTHER BOTHE

«for the coincidence method and his discoveries made therewith»

Physics 1954

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

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

Research into the laws valid for the movement of the electrons around the nucleus in the centre of the atom has been a central problem for physics during this century. Niels Bohr made a start on the solution to the problem in 1913. But his theory was of a provisional nature. Professor Max Born took an active part in striving to improve it, as did the many followers who gathered round him in Göttingen. During the twenties of this century, Göttingen, together with Copenhagen and Munich, was a place of pilgrimage for researchers in the field of atomic theory. When the young Heisenberg, formerly a pupil of Sommerfeld in Munich and of Bohr in Copenhagen, published his epoch-making preliminary work on the exact laws for atomic phenomena in 1925, he was Born's assistant in Göttingen. His work was immediately continued by Born, who gave logical mathematical form to the Heisenberg theory. Owing to this progress, Born, in collaboration with his pupil Jordan and later with Heisenberg also, was able to expand the latter's original results into a comprehensive theory for atomic phenomena. This theory was called *quantum mechanics*.

The following year Born got a new result of fundamental significance. Schrödinger had just then found a new formulation for quantum mechanics. Schrödinger's work expanded the earlier ideas of De Broglie which imply that atomic phenomena are connected with a wave undulation. However, Schrödinger had not solved the problem of how it is possible to make statements about the positions and velocities of particles if one knows the wave corresponding to the particle.

Born provided the solution to the problem. He found that the waves determine the probability of the measuring results. For this reason, according to Born, quantum mechanics gives only a statistical description. This can be illustrated by a simple example. When you shoot at a target it is possible in principle – according to the older conception – to aim the shot from the start so that it is certain to hit the target in the middle. Quantum mechanics teaches us to the contrary – that in principle we cannot predict where a single

shot will hit the target. But we can achieve this much, that from a large number of shots the average point of impact will lie in the middle of the target. In contradiction to the deterministic predictions of the older mechanics, quantum mechanics accordingly poses laws which are of a statistical character, and as regards single phenomena will only determine the probabilities that one or another of various possibilities will occur. For material bodies of ordinary dimensions the uncertainty of the predictions of quantum mechanics is practically of no significance. But in atomic phenomena, on the other hand, it is fundamental. Such a radical break with older ideas could not of course prevail without opposition. But Born's conception is now generally accepted by physicists, with a few exceptions.

In addition to these achievements, which have been rewarded with the Nobel Prize, Born has made fundamental contributions to many fields of physics. In the first place he dedicated his interest to the theory of crystals and has been one of the great pioneers in that field.

After Born had left Göttingen in 1933 he continued his famous researches in Britain, especially as a Professor at Edinburgh.

Professor Walther Bothe, who shares this year's Nobel Prize with Professor Born, began his scientific activity as a theoretical physicist.

The work for which he has now been rewarded with the Nobel Prize was carried out by him in Berlin actually as an experimental physicist. These labours were based on a new use of counter tubes. A counter tube has the property of transmitting an electric current when a charged particle, e.g. an electron, passes through it; and also, with special contrivances, when a light particle collides with it. Bothe's idea was to use two counter tubes in such a manner that the two tubes would only register simultaneous collisions. Such coincidences can only come from two particles emitted in the same elemental process, or from a particle which has travelled through both tubes at high velocity so that the time it takes for the particle's passing from one tube to the other can be neglected.

Bothe used this coincidence method in 1925 and also with improved apparatus about ten years later in order to decide whether the energy rule as well as its complement, the so-called impulse rule, is valid for every collision between a light particle and an electron – as Einstein and Compton assumed – or whether those rules are valid only on average for a large number of collisions – as Bohr and his collaborators had inferred. By investigating light particles and electrons by the coincidence method, Bothe and his co-workers

were able to show convincingly that the rules mentioned are valid for every individual collision. This result was of great significance in principle. The coincidence method has been widely used in the study of cosmic radiation and is one of the most important experimental aids in the investigation of cosmic radiation. This method was first used in this way by Bothe when he was working with Kolhörster who had already given important contributions in the field of cosmic radiation. Bothe and Kolhörster used the coincidence method to pick out those particles in the cosmic radiation which had travelled through two counter tubes. The absorption of cosmic radiation into various materials was determined by placing layers of these substances between the tubes and studying the corresponding reduction in the number of coincidences. It was found that these particles are absorbed at about the same extent as the total cosmic radiation. From the experiments the particularly important result was obtained, that at sea level, cosmic radiation consists in the main of particles of very high penetration.

Bothe and other researchers later improved the coincidence method and extended its field of application. This method has now become one of the most important aids in the study of both nuclear reactions and cosmic radiation.

By many other discoveries and penetrating investigations also, Bothe has enriched our knowledge in these fields in very great measure and has provided an important stimulus to other researchers.

MAX BORN

The statistical interpretation of quantum mechanics

Nobel Lecture, December 11, 1954

The work, for which I have had the honour to be awarded the Nobel Prize for 1954, contains no discovery of a fresh natural phenomenon, but rather the basis for a new mode of thought in regard to natural phenomena. This way of thinking has permeated both experimental and theoretical physics to such a degree that it hardly seems possible to say anything more about it that has not been already so often said. However, there are some particular aspects which I should like to discuss on what is, for me, such a festive occasion. The first point is this: the work at the Göttingen school, which I directed at that time (1926–1927), contributed to the solution of an intellectual crisis into which our science had fallen as a result of Planck's discovery of the quantum of action in 1900. Today, physics finds itself in a similar crisis – I do not mean here its entanglement in politics and economics as a result of the mastery of a new and frightful force of Nature, but I am considering more the logical and epistemological problems posed by nuclear physics. Perhaps it is well at such a time to recall what took place earlier in a similar situation, especially as these events are not without a definite dramatic flavour.

The second point I wish to make is that when I say that the physicists had accepted the concepts and mode of thought developed by us at the time, I am not quite correct. There are some very noteworthy exceptions, particularly among the very workers who have contributed most to building up the quantum theory. Planck, himself, belonged to the sceptics until he died. Einstein, De Broglie, and Schrödinger have unceasingly stressed the unsatisfactory features of quantum mechanics and called for a return to the concepts of classical, Newtonian physics while proposing ways in which this could be done without contradicting experimental facts. Such weighty views cannot be ignored. Niels Bohr has gone to a great deal of trouble to refute the objections. I, too, have ruminated upon them and believe I can make some contribution to the clarification of the position. The matter concerns the borderland between physics and philosophy, and so my physics lecture

will partake of both history and philosophy, for which I must crave your indulgence.

First of all, I will explain how quantum mechanics and its statistical interpretation arose. At the beginning of the twenties, every physicist, I think, was convinced that Planck's quantum hypothesis was correct. According to this theory *energy* appears in finite quanta of magnitude $h\nu$ in oscillatory processes having a specific frequency ν (e.g. in light waves). Countless experiments could be explained in this way and always gave the same value of Planck's constant h. Again, Einstein's assertion that light quanta have *momentum* $h\nu/c$ (where c is the speed of light) was well supported by experiment (e.g. through the Compton effect). This implied a revival of the corpuscular theory of light for a certain complex of phenomena. The wave theory still held good for other processes. Physicists grew accustomed to this *duality* and learned how to cope with it to a certain extent.

In 1913 Niels Bohr had solved the riddle of *line spectra* by means of the quantum theory and had thereby explained broadly the amazing stability of the atoms, the structure of their electronic shells, and the Periodic System of the elements. For what was to come later, the most important assumption of his teaching was this: an atomic system cannot exist in all mechanically possible states, forming a continuum, but in a series of discrete «stationary» states. In a transition from one to another, the difference in energy $E_m - E_n$ is emitted or absorbed as a light quantum $h\nu_{mn}$ (according to whether E_m is greater or less than E_n). This is an interpretation in terms of energy of the fundamental law of spectroscopy discovered some years before by W. Ritz. The situation can be taken in at a glance by writing the energy levels of the stationary states twice over, horizontally and vertically. This produces a square array

	E_1	E_2	E_3	$\cdots$
E_1	11	12	13	—
E_2	21	22	23	—
E_3	31	32	33	—
$\vdots$	—	—	—	—

in which positions on a diagonal correspond to states, and non-diagonal positions correspond to transitions.

It was completely clear to Bohr that the law thus formulated is in conflict with mechanics, and that therefore the use of the energy concept in this

connection is problematical. He based this daring fusion of old and new on his *principle of correspondence*. This consists in the obvious requirement that ordinary classical mechanics must hold to a high degree of approximation in the limiting case where the numbers of the stationary states, the so-called quantum numbers, are very large (that is to say, far to the right and to the lower part in the above array) and the energy changes relatively little from place to place, in fact practically continuously.

Theoretical physics maintained itself on this concept for the next ten years. The problem was this: an harmonic oscillation not only has a frequency, but also an intensity. For each transition in the array there must be a corresponding intensity. The question is how to find this through the considerations of correspondence? It meant guessing the unknown from the available information on a known limiting case. Considerable success was attained by Bohr himself, by Kramers, Sommerfeld, Epstein, and many others. But the decisive step was again taken by Einstein who, by a fresh derivation of Planck's radiation formula, made it transparently clear that the classical concept of intensity of radiation must be replaced by the statistical concept of *transition probability*. To each place in our pattern or array there belongs (together with the frequency $\nu_{mn} = (E_n - E_m)/h$) a definite probability for the transition coupled with emission or absorption.

In Göttingen we also took part in efforts to distil the unknown mechanics of the atom from the experimental results. The logical difficulty became ever sharper. Investigations into the scattering and dispersion of light showed that Einstein's conception of transition probability as a measure of the strength of an oscillation did not meet the case, and the idea of an *amplitude* of oscillation associated with each transition was indispensable. In this connection, work by Ladenburg[1], Kramers[2], Heisenberg[3], Jordan and me[4] should be mentioned. The art of guessing correct formulae, which deviate from the classical formulae, yet contain them as a limiting case according to the correspondence principle, was brought to a high degree of perfection. A paper of mine, which introduced, for the first time I think, the expression *quantum mechanics* in its title, contains a rather involved formula (still valid today) for the reciprocal disturbance of atomic systems.

Heisenberg, who at that time was my assistant, brought this period to a sudden end[5]. He cut the Gordian knot by means of a philosophical principle and replaced guess-work by a mathematical rule. The principle states that concepts and representations that do not correspond to physically observable facts are not to be used in theoretical description. Einstein used the

same principle when, in setting up his theory of relativity, he eliminated the concepts of absolute velocity of a body and of absolute simultaneity of two events at different places. Heisenberg banished the picture of electron orbits with definite radii and periods of rotation because these quantities are not observable, and insisted that the theory be built up by means of the square arrays mentioned above. Instead of describing the motion by giving a co-ordinate as a function of time, $x(t)$, an array of transition amplitudes x_{mn} should be determined. To me the decisive part of his work is the demand to determine a rule by which from a given

$$\text{array} \begin{bmatrix} x_{11}\ x_{12} \cdots \cdots \\ x_{21}\ x_{22} \cdots \cdots \\ - - - - - - - \end{bmatrix} \text{ the array for the square } \begin{bmatrix} (x^2)_{11}\ (x^2)_{12} \cdots \\ (x^2)_{21}\ (x^2)_{22} \cdots \\ - - - - - - - - \end{bmatrix}$$

can be found (or, more general, the *multiplication rule* for such arrays).

By observation of known examples solved by guess-work he found this rule and applied it successfully to simple examples such as the harmonic and anharmonic oscillator.

This was in the summer of 1925. Heisenberg, plagued by hay fever took leave for a course of treatment by the sea and gave me his paper for publication if I thought I could do something with it.

The significance of the idea was at once clear to me and I sent the manuscript to the *Zeitschrift für Physik*. I could not take my mind off Heisenberg's multiplication rule, and after a week of intensive thought and trial I suddenly remembered an algebraic theory which I had learned from my teacher, Professor Rosanes, in Breslau. Such square arrays are well known to mathematicians and, in conjunction with a specific rule for multiplication, are called matrices. I applied this rule to Heisenberg's quantum condition and found that this agreed in the diagonal terms. It was easy to guess what the remaining quantities must be, namely, zero; and at once there stood before me the peculiar formula

$$pq - qp = h/2\pi i$$

This meant that coordinates q and momenta p cannot be represented by figure values but by symbols, the product of which depends upon the order of multiplication – they are said to be «non-commuting».

I was as excited by this result as a sailor would be who, after a long voyage, sees from afar, the longed-for land, and I felt regret that Heisenberg was not

there. I was convinced from the start that we had stumbled on the right path. Even so, a great part was only guess-work, in particular, the disappearance of the non-diagonal elements in the above-mentioned expression. For help in this problem I obtained the assistance and collaboration of my pupil Pascual Jordan, and in a few days we were able to demonstrate that I had guessed correctly. The joint paper by Jordan and myself[6] contains the most important principles of quantum mechanics including its extension to electrodynamics. There followed a hectic period of collaboration among the three of us, complicated by Heisenberg's absence. There was a lively exchange of letters; my contribution to these, unfortunately, have been lost in the political disorders. The result was a three-author paper[7] which brought the formal side of the investigation to a definite conclusion. Before this paper appeared, came the *first dramatic surprise*: Paul Dirac's paper on the same subject[8]. The inspiration afforded by a lecture of Heisenberg's in Cambridge had led him to similar results as we had obtained in Göttingen except that he did not resort to the known matrix theory of the mathematicians, but discovered the tool for himself and worked out the theory of such non-commutating symbols.

The first non-trivial and physically important application of quantum mechanics was made shortly afterwards by W. Pauli[9] who calculated the stationary energy values of the *hydrogen atom* by means of the matrix method and found complete agreement with Bohr's formulae. From this moment onwards there could no longer be any doubt about the correctness of the theory.

What this formalism really signified was, however, by no means clear. Mathematics, as often happens, was cleverer than interpretative thought. While we were still discussing this point there came the *second dramatic surprise*, the appearance of Schrödinger's famous papers[10]. He took up quite a different line of thought which had originated from Louis de Broglie[11].

A few years previously, the latter had made the bold assertion, supported by brilliant theoretical considerations, that wave-corpuscle duality, familiar to physicists in the case of light, must also be valid for electrons. To each electron moving free of force belongs a plane wave of a definite wavelength which is determined by Planck's constant and the mass. This exciting dissertation by De Broglie was well known to us in Göttingen. One day in 1925 I received a letter from C. J. Davisson giving some peculiar results on the reflection of electrons from metallic surfaces. I, and my colleague on the experimental side, James Franck, at once suspected that these curves of

Davisson's were crystal-lattice spectra of De Broglie's electron waves, and we made one of our pupils, Elsasser[12], to investigate the matter. His result provided the first preliminary confirmation of the idea of De Broglie's, and this was later proved independently by Davisson and Germer[13] and G. P. Thomson[14] by systematic experiments.

But this acquaintance with De Broglie's way of thinking did not lead us to an attempt to apply it to the electronic structure in atoms. This was left to Schrödinger. He extended De Broglie's wave equation which referred to force-free motion, to the case where the effect of force is taken into account, and gave an exact formulation of the *subsidiary conditions*, already suggested by De Broglie, to which the wave function ψ must be subjected, namely that it should be single-valued and finite in space and time. And he was successful in deriving the stationary states of the hydrogen atom in the form of those monochromatic solutions of his wave equation which do not extend to infinity.

For a brief period at the beginning of 1926, it looked as though there were, suddenly, two self-contained but quite distinct systems of explanation extant: matrix mechanics and wave mechanics. But Schrödinger himself soon demonstrated their complete equivalence.

Wave mechanics enjoyed a very great deal more popularity than the Göttingen or Cambridge version of quantum mechanics. It operates with a wave function ψ, which in the case of *one* particle at least, can be pictured in space, and it uses the mathematical methods of partial differential equations which are in current use by physicists. Schrödinger thought that his wave theory made it possible to return to deterministic classical physics. He proposed (and he has recently emphasized his proposal anew[15]), to dispense with the particle representation entirely, and instead of speaking of electrons as particles, to consider them as a continuous density distribution $|\psi|^2$ (or electric density $e|\psi|^2$).

To us in Göttingen this interpretation seemed unacceptable in face of well established experimental facts. At that time it was already possible to count particles by means of scintillations or with a Geiger counter, and to photograph their tracks with the aid of a Wilson cloud chamber.

It appeared to me that it was not possible to obtain a clear interpretation of the ψ-function, by considering bound electrons. I had therefore, as early as the end of 1925, made an attempt to extend the matrix method, which obviously only covered oscillatory processes, in such a way as to be applicable to aperiodic processes. I was at that time a guest of the Mas-

sachusetts Institute of Technology in the USA, and I found there in Norbert Wiener an excellent collaborator. In our joint paper[16] we replaced the matrix by the general concept of an operator, and thus made it possible to describe aperiodic processes. Nevertheless we missed the correct approach. This was left to Schrödinger, and I immediately took up his method since it held promise of leading to an interpretation of the ψ-function. Again an idea of Einstein's gave me the lead. He had tried to make the duality of particles – light quanta or photons – and waves comprehensible by interpreting the square of the optical wave amplitudes as probability density for the occurrence of photons. This concept could at once be carried over to the ψ-function: $|\psi|^2$ ought to represent the probability density for electrons (or other particles). It was easy to assert this, but how could it be proved?

The atomic collision processes suggested themselves at this point. A swarm of electrons coming from infinity, represented by an incident wave of known intensity (i.e., $|\psi|^2$), impinges upon an obstacle, say a heavy atom. In the same way that a water wave produced by a steamer causes secondary circular waves in striking a pile, the incident electron wave is partially transformed into a secondary spherical wave whose amplitude of oscillation ψ differs for different directions. The square of the amplitude of this wave at a great distance from the scattering centre determines the relative probability of scattering as a function of direction. Moreover, if the scattering atom itself is capable of existing in different stationary states, then Schrödinger's wave equation gives automatically the probability of excitation of these states, the electron being scattered with loss of energy, that is to say, inelastically, as it is called. In this way it was possible to get a theoretical basis[17] for the assumptions of Bohr's theory which had been experimentally confirmed by Franck and Hertz. Soon Wentzel[18] succeeded in deriving Rutherford's famous formula for the scattering of α-particles from my theory.

However, a paper by Heisenberg[19], containing his celebrated uncertainty relationship, contributed more than the above-mentioned successes to the swift acceptance of the statistical interpretation of the ψ-function. It was through this paper that the revolutionary character of the new conception

e determinism of classical physics
ncept of reality which looked upon
were very small grains of sand. At
e position and velocity. This is not
is determined with increasing ac-
e velocity becomes less and *vice*

became clear. It showed that not only
must be abandonded, but also the naiv
the particles of atomic physics as if th
every instant a grain of sand has a def
the case with an electron. If its positi
curacy, the possibility of ascertainin

versa. I shall return shortly to these problems in a more general connection, but would first like to say a few words about the theory of collisions.

The mathematical approximation methods which I used were quite primitive and soon improved upon. From the literature, which has grown to a point where I cannot cope with, I would like to mention only a few of the first authors to whom the theory owes great progress: Faxén in Sweden, Holtsmark in Norway[20], Bethe in Germany[21], Mott and Massey in England[22].

Today, collision theory is a special science with its own big, solid textbooks which have grown completely over my head. Of course in the last resort all the modern branches of physics, quantum electrodynamics, the theory of mesons, nuclei, cosmic rays, elementary particles and their transformations, all come within range of these ideas and no bounds could be set to a discussion on them.

I should also like to mention that in 1926 and 1927 I tried another way of supporting the statistical concept of quantum mechanics, partly in collaboration with the Russian physicist Fock[23]. In the above-mentioned three-author paper there is a chapter which anticipates the Schrödinger function, except that it is not thought of as a function $\psi(x)$ in space, but as a function ψ_n of the discrete index $n = 1, 2, \ldots$ which enumerates the stationary states. If the system under consideration is subject to a force which is variable with time, ψ_n becomes also time-dependent, and $|\psi_n(t)|^2$ signifies the probability for the existence of the state n at time t. Starting from an initial distribution where there is only one state, transition probabilities are obtained, and their properties can be examined. What interested me in particular at the time, was what occurs in the adiabatic limiting case, that is, for very slowly changing action. It was possible to show that, as could have been expected, the probability of transitions becomes ever smaller. The theory of transition probabilities was developed independently by Dirac with great success. It can be said that the whole of atomic and nuclear physics works with this system of concepts, particularly in the very elegant form given to them by Dirac[24]. Almost all experiments lead to statements about relative frequencies of events, even when they occur concealed under such names as effective cross section or the like.

How does it come about then, that great scientists such as Einstein, Schrödinger, and De Broglie are nevertheless dissatisfied with the situation? Of course, all these objections are levelled not against the correctness of the formulae, but against their interpretation. Two closely knitted points of view

are to be distinguished: the question of determinism and the question of reality.

Newtonian mechanics is deterministic in the following sense:

If the initial state (positions and velocities of all particles) of a system is accurately given, then the state at any other time (earlier or later) can be calculated from the laws of mechanics. All the other branches of classical physics have been built up according to this model. Mechanical determinism gradually became a kind of article of faith: the world as a machine, an automaton. As far as I can see, this idea has no forerunners in ancient and medieval philosophy. The idea is a product of the immense success of Newtonian mechanics, particularly in astronomy. In the 19th century it became a basic philosophical principle for the whole of exact science. I asked myself whether this was really justified. Can absolute predictions really be made for all time on the basis of the classical equations of motion? It can easily be seen, by simple examples, that this is only the case when the possibility of absolutely exact measurement (of position, velocity, or other quantities) is assumed. Let us think of a particle moving without friction on a straight line between two end-points (walls), at which it experiences completely elastic recoil. It moves with constant speed equal to its initial speed v_o backwards and forwards, and it can be stated exactly where it will be at a given time provided that v_o is accurately known. But if a small inaccuracy Δv_o is allowed, then the inaccuracy of prediction of the position at time t is $t\Delta v_o$ which increases with t. If one waits long enough until time $t_c = l/\Delta v_o$ where l is the distance between the elastic walls, the inaccuracy Δx will have become equal to the whole space l. Thus it is impossible to forecast anything about the position at a time which is later than t_c. Thus determinism lapses completely into indeterminism as soon as the slightest inaccuracy in the data on velocity is permitted. Is there any sense – and I mean any physical sense, not metaphysical sense – in which one can speak of absolute data? Is one justified in saying that the coordinate $x = \pi$ cm where $\pi = 3.1415..$ is the familiar transcendental number that determines the ratio of the circumference of a circle to its diameter? As a mathematical tool the concept of a real number represented by a nonterminating decimal fraction is exceptionally important and fruitful. As the measure of a physical quantity it is nonsense. If π is taken to the 20th or the 25th place of decimals, two numbers are obtained which are indistinguishable from each other and the true value of π by any measurement. According to the heuristic principle used by Einstein in the theory of relativity, and by Heisenberg in the quantum theory, concepts which correspond to no conceivable observation should be eliminated

from physics. This is possible without difficulty in the present case also. It is only necessary to replace statements like $x = \pi$ cm by: the probability of distribution of values of x has a sharp maximum at $x = \pi$ cm; and (if it is desired to be more accurate) to add: of such and such a breadth. In short, ordinary mechanics must also be statistically formulated. I have occupied myself with this problem a little recently, and have realized that it is possible without difficulty. This is not the place to go into the matter more deeply. I should like only to say this: the determinism of classical physics turns out to be an illusion, created by overrating mathematico-logical concepts. It is an idol, not an ideal in scientific research and cannot, therefore, be used as an objection to the essentially indeterministic statistical interpretation of quantum mechanics.

Much more difficult is the objection based on reality. The concept of a particle, e.g. a grain of sand, implicitly contains the idea that it is in a definite position and has definite motion. But according to quantum mechanics it is impossible to determine simultaneously with any desired accuracy both position and velocity (more precisely: momentum, i.e. mass times velocity). Thus two questions arise: what prevents us, in spite of the theoretical assertion, to measure both quantities to any desired degree of accuracy by refined experiments? Secondly, if it really transpires that this is not feasible, are we still justified in applying to the electron the concept of particle and therefore the ideas associated with it?

Referring to the first question, it is clear that if the theory is correct – and we have ample grounds for believing this – the obstacle to simultaneous measurement of position and motion (and of other such pairs of so-called conjugate quantities) must lie in the laws of quantum mechanics themselves. In fact, this is so. But it is not a simple matter to clarify the situation. Niels Bohr himself has gone to great trouble and ingenuity[25] to develop a theory of measurements to clear the matter up and to meet the most refined and ingenious attacks of Einstein, who repeatedly tried to think out methods of measurement by means of which position and motion could be measured simultaneously and accurately. The following emerges: to measure space coordinates and instants of time, rigid measuring rods and clocks are required. On the other hand, to measure momenta and energies, devices are necessary with movable parts to absorb the impact of the test object and to indicate the size of its momentum. Paying regard to the fact that quantum mechanics is competent for dealing with the interaction of object and apparatus, it is seen that no arrangement is possible that will fulfil both require-

ments simultaneously. There exist, therefore, mutually exclusive though complementary experiments which only as a whole embrace everything which can be experienced with regard to an object.

This idea of *complementarity* is now regarded by most physicists as the key to the clear understanding of quantum processes. Bohr has generalized the idea to quite different fields of knowledge, e.g. the connection between consciousness and the brain, to the problem of free will, and other basic problems of philosophy. To come now to the last point: can we call something with which the concepts of position and motion cannot be associated in the usual way, a thing, or a particle? And if not, what is the reality which our theory has been invented to describe?

The answer to this is no longer physics, but philosophy, and to deal with it thoroughly would mean going far beyond the bounds of this lecture. I have given my views on it elsewhere[26]. Here I will only say that I am emphatically in favour of the retention of the particle idea. Naturally, it is necessary to redefine what is meant. For this, well-developed concepts are available which appear in mathematics under the name of invariants in transformations. Every object that we perceive appears in innumerable aspects. The concept of the object is the invariant of all these aspects. From this point of view, the present universally used system of concepts in which particles and waves appear simultaneously, can be completely justified.

The latest research on nuclei and elementary particles has led us, however, to limits beyond which this system of concepts itself does not appear to suffice. The lesson to be learned from what I have told of the origin of quantum mechanics is that probable refinements of mathematical methods will not suffice to produce a satisfactory theory, but that somewhere in our doctrine is hidden a concept, unjustified by experience, which we must eliminate to open up the road.

1. R. Ladenburg, *Z. Physik*, 4 (1921) 451; R. Ladenburg and F. Reiche, *Naturwiss.*, 11 (1923) 584.
2. H. A. Kramers, *Nature*, 113 (1924) 673.
3. H. A. Kramers and W. Heisenberg, *Z. Physik*, 31 (1925) 681.
4. M. Born, *Z. Physik*, 26 (1924) 379; M. Born and P. Jordan, *Z. Physik*, 33 (1925) 479.
5. W. Heisenberg, *Z. Physik*, 33 (1925) 879.
6. M. Born and P. Jordan, *Z. Physik*, 34 (1925) 858.

7. M. Born, W. Heisenberg, and P. Jordan, *Z. Physik*, 35 (1926) 557.
8. P. A. M. Dirac, *Proc. Roy. Soc. (London)*, A 109 (1925) 642.
9. W. Pauli, *Z. Physik*, 36 (1926) 336.
10. E. Schrödinger, *Ann. Physik*, [4] 79 (1926) 361, 489, 734; 80 (1926) 437; 81 (1926) 109.
11. L. de Broglie, *Thesis Paris*, 1924; *Ann. Phys. (Paris)*, [10] 3 (1925) 22.
12. W. Elsasser, *Naturwiss.*, 13 (1925) 711.
13. C. J. Davisson and L. H. Germer, *Phys. Rev.*, 30 (1927) 707.
14. G. P. Thomson and A. Reid, *Nature*, 119 (1927) 890; G. P. Thomson, *Proc. Roy. Soc. (London)*, A 117 (1928) 600.
15. E. Schrödinger, *Brit. J. Phil. Sci.*, 3 (1952) 109, 233.
16. M. Born and N. Wiener, *Z. Physik*, 36 (1926) 174.
17. M. Born, *Z. Physik*, 37 (1926) 863; 38 (1926) 803; *Göttinger Nachr. Math. Phys. Kl.*, (1926) 146.
18. G. Wentzel, *Z. Physik*, 40 (1926) 590.
19. W. Heisenberg, *Z. Physik*, 43 (1927) 172.
20. H. Faxén and J. Holtsmark, *Z. Physik*, 45 (1927) 307.
21. H. Bethe, *Ann. Physik*, 5 (1930) 325.
22. N. F. Mott, *Proc. Roy. Soc. (London)*, A 124 (1929) 422, 425; *Proc. Cambridge Phil. Soc.*, 25 (1929) 304.
23. M. Born, *Z. Physik*, 40 (1926) 167; M. Born and V. Fock, *Z. Physik*, 51 (1928) 165.
24. P. A. M. Dirac, *Proc. Roy. Soc. (London)*, A 109 (1925) 642; 110 (1926) 561; 111 (1926) 281; 112 (1926) 674.
25. N. Bohr, *Naturwiss.*, 16 (1928) 245; 17 (1929) 483; 21 (1933) 13. «Kausalität und Komplementarität» (Causality and Complementarity), *Die Erkenntnis*, 6 (1936) 293.
26. M. Born, *Phil. Quart.*, 3 (1953) 134; *Physik. Bl.*, 10 (1954) 49.

Biography

Max Born was born in Breslau on the 11th December, 1882, to Professor Gustav Born, anatomist and embryologist, and his wife Margarete, *née* Kauffmann, who was a member of a Silesian family of industrialists.

Max attended the König Wilhelm's Gymnasium in Breslau and continued his studies at the Universities of Breslau (where the well-known mathematician Rosanes introduced him to matrix calculus), Heidelberg, Zurich (here he was deeply impressed by Hurwitz's lectures on higher analysis), and Göttingen. In the latter seat of learning he read mathematics chiefly, sitting under Klein, Hilbert, Minkowski, and Runge, but also studied astronomy under Schwarzschild, and physics under Voigt. He was awarded the Prize of the Philosophical Faculty of the University of Göttingen for his work on the stability of elastic wires and tapes in 1906, and graduated at this university a year later on the basis of this work.

Born next went to Cambridge for a short time, to study under Larmor and J. J. Thomson. Back in Breslau during the years 1908–1909, he worked with the physicists Lummer and Pringsheim, and also studied the theory of relativity. On the strength of one of his papers, Minkowski invited his collaboration at Göttingen but soon after his return there, in the winter of 1909, Minkowski died. He had then the task of sifting Minkowski's literary works in the field of physics and of publishing some uncompleted papers. Soon he became an academic lecturer at Göttingen in recognition of his work on the relativistic electron. He accepted Michelson's invitation to lecture on relativity in Chicago (1912) and while there he did some experiments with the Michelson grating spectrograph.

An appointment as professor (extraordinarius) to assist Max Planck at Berlin University came to Born in 1915 but he had to join the German Armed Forces. In a scientific office of the army he worked on the theory of sound ranging. He found time also to study the theory of crystals, and published his first book, *Dynamik der Kristallgitter* (Dynamics of Crystal Lattices), which summarized a series of investigations he had started at Göttingen.

At the conclusion of the First World War, in 1919, Born was appointed Professor at the University of Frankfurt-on-Main, where a laboratory was put at his disposal. His assistant was Otto Stern, and the first of the latter's well-known experiments, which later were rewarded with a Nobel Prize, originated there.

Max Born went to Göttingen as Professor in 1921, at the same time as James Franck, and he remained there for twelve years, interrupted only by a trip to America in 1925. During these years the Professor's most important works were created; first a modernized version of his book on crystals, and numerous investigations by him and his pupils on crystal lattices, followed by a series of studies on the quantum theory. Among his collaborators at this time were many physicists, later to become well-known, such as Pauli, Heisenberg, Jordan, Fermi, Dirac, Hund, Hylleraas, Weisskopf, Oppenheimer, Joseph Mayer and Maria Goeppert-Mayer. During the years 1925 and 1926 he published, with Heisenberg and Jordan, investigations on the principles of quantum mechanics (matrix mechanics) and soon after this, his own studies on the statistical interpretation of quantum mechanics.

As were so many other German scientists, he was forced to emigrate in 1933 and was invited to Cambridge, where he taught for three years as Stokes Lecturer. His main sphere of work during this period was in the field of nonlinear electrodynamics, which he developed in collaboration with Infeld.

During the winter of 1935–1936 Born spent six months in Bangalore at the Indian Institute of Science, where he worked with Sir C. V. Raman and his pupils. In 1936 he was appointed Tait Professor of Natural Philosophy in Edinburgh, where he worked until his retirement in 1953. He is now living at the small spa town, Bad Pyrmont.

Max Born has been awarded fellowships of many academies – Göttingen, Moscow, Berlin, Bangalore, Bucharest, Edinburgh, London, Lima, Dublin, Copenhagen, Stockholm, Washington, and Boston, and he has received honorary doctorates from Bristol, Bordeaux, Oxford, Freiburg/Breisgau, Edinburgh, Oslo, Brussels Universities, Humboldt University Berlin, and Technical University Stuttgart. He holds the Stokes Medal of Cambridge, the Max Planck Medaille der Deutschen Physikalischen Gesellschaft (i.e. of the German Physical Society); the Hughes Medal of the Royal Society, London, the Hugo Grotius Medal for International Law, and was also awarded the MacDougall-Brisbane Prize and the Gunning-Victoria Jubilee Prize of the Royal Society, Edinburgh. In 1953 he was made honorary citizen of

the town of Göttingen and a year later was granted the Nobel Prize for Physics. He was awarded the Grand Cross of Merit with Star of the Order of Merit of the German Federal Republic in 1959.

The year 1913 saw his marriage to Hedwig, *née* Ehrenberg, and there are three children of the marriage.

WALTHER BOTHE

The coincidence method

*Nobel Lecture**

Before embarking on the subject of my lecture, permit me to devote a few words to the man to whom, apart from my teacher, Max Planck, I owe so much, and who died ten years ago after a long period of painful suffering. In 1912 Hans Geiger was appointed Director of a new Laboratory for Radioactivity at the Physikalisch-Technische Reichsanstalt, Berlin-Charlottenburg of which Emil Warburg was then the President; previous to this, he had worked for six years under Rutherford at Manchester. In June 1913, I became Geiger's assistant. The Laboratory for Radioactivity consisted of only two rooms at the time; at a later date, when tests of radioactive substances became more extensive, it expanded into four rooms. This modesty of his room requirements – Geiger repeatedly stated that he had no desire for a giant institute – is characteristic of the principal trait in Geiger's personality as a scientist: the desire to keep scientific work within economic bounds. No doubt, the unique influence of Rutherford had something to do with this; equally indubitably, this influence harmonized with a natural tendency. However this may be, the experiments by Geiger and Marsden on the scattering of alpha rays are known to form part of the beginning of the entire experimental atom physics of recent days. I think the main lesson which I have learnt from Geiger is to select from a large number of possible and perhaps useful experiments that which appears the most urgent at the moment, and to do this experiment with the simplest possible apparatus, i.e. clearly arranged and variable apparatus.

It was in 1924 that I came across the theoretical paper by Bohr, Kramers, and Slater, which had just been published and which suggested a possible interpretation of the wave–particle dualism in the accepted description of the properties of light. This must be understood to mean the experimental fact that light of all wavelengths behaves as a wave process (interference) with pure propagation, but behaves as particles (light quanta: photo-effect, Compton effect) on conversion into other types of energy. The new idea

* Owing to Professor Bothe's illness the lecture was not given orally.

consisted in denying strict validity to the energy-impulse law. In the individual or elementary process, so long as only a single act of emission was involved, the laws of conservation were held to be statistically satisfied only, to become valid for a macroscopic totality of a very large number of elementary processes only, so that there was no conflict with the available empirical evidence. It was immediately obvious that this question would have to be decided experimentally, before definite progress could be made. That such a decision was *possible*, Geiger and I agreed immediately, when I discussed the paper by Bohr, Kramers, and Slater with Geiger.

The experimental problem offered several means of attack. We decided in favour of an experiment with the effect discovered a short time previously by A. H. Compton, i.e. the scattering of light on practically free electrons. Apart from the scattered light, there occur the «recoil electrons» which had been observed and interpreted by C. T. R. Wilson in the cloud chamber, and by me both in the cloud chamber and by an ionization method. The «question to Nature» which the experiment was designed to answer could therefore be formulated as follows: is it exactly a scatter quantum and a recoil electron that are simultaneously emitted in the elementary process, or is there merely a statistical relationship between the two?

Meanwhile, Geiger had developed the so-called needle counter which has the advantage of responding not only to heavy particles but also to electrons, and therefore to light quanta of sufficiently high energy capable of releasing electrons within the counter.

Our arrangement therefore consisted of two needle counters, past the common front wall of which, without touching it, swept a beam of X-rays. The X-ray beam travelled in a hydrogen atmosphere; the Compton processes occurred in the one counter which indicated the recoil electrons, whereas only the scatter quanta were able to penetrate into the other counter and actuated it by electron release with very much lower probability. The readings of both counters were recorded side by side on a moving paper chart. In this way we succeeded after a few failures to establish the accuracy of any temporal «coincidence» between the two pointer readings as being 10^{-4} sec. Film consumption however was so enormous that our laboratory with the film strips strung up for drying sometimes resembled an industrial laundry.

The final result we obtained was that systematic coincidences do indeed occur with the frequency that could be estimated from the experimental geometry and the response probabilities of the counters on the assumption

that, in each elementary Compton process, a scatter quantum and a recoil electron are generated *simultaneously*. The strict validity of the law of the conservation of energy even in the elementary process had been demonstrated, and the ingenious way out of the wave–particle problem discussed by Bohr, Kramers, and Slater was shown to be a blind alley.

This result was confirmed by different researchers using various experimental arrangements. When, more than ten years later, some doubts as to the correctness of this result were voiced, I tried with my then assistant, H. Maier-Leibnitz, to supplement and improve the original experiment in one point: the object was to demonstrate both simultaneity and uniformity of direction of scatter quantum and recoil electron, as was to be expected according to Compton's theory, i.e. according to the laws of elastic impact between two bodies. On this occasion, we employed the energy-rich gamma radiation of a radiothorium preparation. Again, the result was clearly positive. This demonstrated both the conservation of energy and the conservation of the impulse.

Unfortunately, the collaboration with Geiger came to an end in 1925, when Geiger was called to Kiel University. When dividing up the field on which we had hitherto worked together, the coincidence method was, at Geiger's generous suggestion, allocated to me.

The possibility of the purely statistical validity of the conservation theorems discussed by Bohr, Kramers, and Slater appeared sufficiently important to be tested in yet another case. A spherical wave is emitted in the elementary process of *light emission*. The problem was: can this spherical wave initiate an absorption act in one direction of emission only, as the energy theorem postulates, or can it do so also statistically independently in several directions, as is to be expected according to Bohr, Kramers, and Slater? It must be borne in mind in an experiment of this kind, that, by contrast with the Compton effect, the probability of demonstrating an absorption act may not be of an order of magnitude much below unity, because otherwise any systematic coincidences that might occur would be submerged in the inevitable accidental coincidences. This was achieved by harmonizing the radiation source (iron or copper-K-fluorescence radiation) and the gas charge of the needle counters (argon) erected on either side so that the absorption probability in the gas charge was as close as possible to unity. Besides, the solid angles which the two counters offered to the radiation source had to amount as far as possible to 2π. The result of this experiment (1926) was that *no* systematic coincidences occurred, at least not with

the frequency to be expected according to Bohr, Kramers, and Slater. Strict conservation of energy in the elementary process had thus been confirmed also by a negative experiment. The wave–particle problem was destined to remain open for a short time only. During this time I had the singular good fortune of being able to discuss the problem constantly with Einstein. Some experiments done at Einstein's suggestion yielded no decisively new result. The (at least formal) solution was provided by wave mechanics; it is contained simply in the assumption that the Schrödinger wave of a system consisting of n particles is a wave in the $3n$-dimensional «configuration space».

An entirely different field in which the coincidence method bore fruit, was that of «cosmic radiation» or «ultra radiation» as its discoverer, Hess, called it. Meanwhile, Geiger had developed, in Kiel, the powerful tool of the Geiger-Müller counter. Coincidences between unscreened counters, caused by cosmic rays, had been observed both by Geiger himself and by W. Kolhörster, then a guest in my Berlin laboratory. More profound discoveries were to be expected by arranging absorbing layers of variable thickness between or/and above the counters. Such experiments which I conducted together with Kolhörster in 1929 prompted the daring conclusion that cosmic radiation does not consist primarily of gamma rays, as had generally been assumed previously because of the high permeating power, but of material particles with an energy of at least 1,000 million electron volt. Such counter-coincidence arrangements were increasingly used in the period which followed, using increasing numbers of counters, in part combined with cloud chambers, ionization chambers, scintillation counters, etc. The material particle nature of primary cosmic radiation has been confirmed, although the processes turned out to be extraordinarily more complicated than we had assumed. As a simple example of this we would only mention that B. Rossi who also spent some time as guest in my PTR laboratory, later succeeded in observing by means of coincidences between *juxtaposed* counters («Rossi curve») the first signs of the occurrence of showers of particles. The possible applications of the coincidence method to the subject of cosmic radiation have by no means been exhausted yet.

The same principle of measurement as in cosmic radiation can of course also be applied to ordinary beta and gamma rays. It is for example possible to determine in a very simple manner, with the assistance of only two counters and a variable absorber between them, the mean gamma energy in a mixture of gamma rays and their secondary electrons (Bothe and Becker,

1930). This method can be useful where for some reason it is impossible to apply the usual spectrometer method with magnetic deviation.

The technology of coincidence counting has been considerably improved meanwhile. Instead of the complicated photographic recording, we have long since passed on to valve circuits in conjunction with mechanical counters, which provides the advantage of greater simplicity and permits reduction of the so-called resolution period by several orders of magnitude, so that the interfering « accidental » coincidences in many cases play no part at all. I used a circuit employing a multiple-grid coincidence valve as early as 1929. Rossi was the first to describe another system working with valves in parallel; it has the advantage that it can easily be extended to coincidences between more than two events, and is therefore predominantly used today. (Recently, Z. Bay and others succeeded, in the U. S. A., in reducing the coincidence resolution period to 10^{-11} sec by means of multipliers.)

A further large field for the application of the coincidence method is that of nuclear reactions. In a joint investigation with my collaborator H. Fränz (1928) and Pose in Halle it was discovered that in the artificial conversion of a nucleus (^{10}B in our case) by alpha rays, there occur several discrete proton groups of different energy. Shortly afterwards (1930) I discovered, with H. Becker, the gamma rays that are generated on bombarding not only boron, but also other elements, with alpha rays. Both these results found a common interpretation. During conversion, the newly formed nucleus is not always immediately in the ground state, but is at times in one of the possible activated states. In this case, the particle formed has correspondingly less energy, whereas the product nucleus passes into the ground state with emission of the quantity of energy saved as gamma radiation. As a rule, this transition occurs in a period of immeasurably short duration, i.e. practically simultaneously with the emission of the new particle. To demonstrate this simultaneity is by no means trivial, because it may for example happen that the product nucleus *always* forms in an activated state at first. This can be decided by coincidence measurements. In this case, even the most energy-rich group of particles that occurs would have to be coupled with gamma radiation, which is not the case if this group belongs to the ground state of the product nucleus. (For the case of « metastable » states of excitation, these arguments must be modified analogously.) Such measurements were first carried out in 1925 by H. J. von Baeyer who was then my student at Heidelberg, again for the case, already mentioned, of boron conversion by alpha rays. In the same manner, it is possible to determine whether two or several

of the gamma quanta generated in a nuclear reaction form in the same nucleus, i.e. practically simultaneously, or whether they are emitted alternatively during the conversion of separate nuclei. Such questions are of importance for the balance of energy, i.e. for the measurement of reaction energies and nuclear mass. Direction coupling between the various radiations generated in a nuclear reaction both with one another and with the initiating radiation can also be detected and measured by coincidences; this provides valuable information about the structure of the atomic nuclei. Analogous problems in spontaneous conversions (natural and artificial radioactivity) can be tackled experimentally in the same manner, as has been demonstrated with RaC decomposition (Bothe and Maier-Leibnitz, 1937).

Many applications of the coincidence method will therefore be found in the large field of nuclear physics, and we can say without exaggeration that the method is one of the essential tools of the modern nuclear physicist.

Biography

Walther Bothe was born on January 8, 1891, at Oranienburg, near Berlin.

From 1908 until 1912 he studied physics at the University of Berlin, where he was a pupil of Max Planck, obtaining his doctorate just before the outbreak of the 1914–1918 war. From 1913 until 1930 he worked at the Physikalisch-Technische Reichsanstalt in the same city, becoming a Professor Extraordinary in the University there. In 1930 he was appointed Professor of Physics, and Director of the Institute of Physics at the University of Giessen.

In 1932 he was appointed Director of the Institute of Physics at the University of Heidelberg, in succession to Philipp Lenard, becoming in 1934 Director of the Institute of Physics at the Max Planck Institute for Medical Research in that city. At the end of the Second World War, when this Institute was taken over for other purposes, Bothe returned to the Department of Physics in the University, where he taught until the illness which had handicapped him for several years compelled him to restrict the scope of his work. He was able, however, to supervise the work of the Institute of Physics in the Max Planck Institute and he continued to do this until his death in Heidelberg on August 2, 1957.

Bothe's scientific work coincided with the opening up of the vast field of nuclear physics and the results he obtained led to new outlooks and methods.

He was, during the First World War, taken prisoner by the Russians and spent a year in captivity in Siberia. This year he devoted to mathematical studies and to learning the Russian language; in 1920 he was sent back to Germany.

He then collaborated with H. Geiger at the Physikalisch-Technische Reichsanstalt in Berlin. Together with Geiger, whose influence determined much of his scientific work, he published, in 1924, his method of coincidence, by which important discoveries were subsequently made. It is based on the fact that, when a single particle passes through two or more Geiger counters, the pulses from each counter are practically coincident in time. The pulse from each counter is then sent to a coincidence circuit which indicates

pulses that are coincident in time. Arrays of Geiger counters in coincidence select particles moving in a given direction and the method can be used, for example, to measure the angular distribution of cosmic rays. Bothe applied this method to the study of the Compton effect and to other problems of physics. Together he and Geiger clarified ideas about the small angle scattering of light rays and Bothe summarized their work on this problem in his *Handbuch* article published in 1926 and 1933, establishing the foundations of modern methods for the analysis of scatter processes. From 1923 until 1926 Bothe concentrated, especially on experimental and theoretical work on the corpuscular theory of light. He had, some months before the discovery of the Compton effect, observed, in a Wilson chamber filled with hydrogen, the short track of the recoil electrons of X-rays and he did further work on the direction of the emission of photo electrons. Together he and Geiger related the Compton effect to the theory of Bohr, Kramers, and Slater, and the results of their work provided strong support for the corpuscular theory of light.

In 1927 Bothe further clarified, by means of his coincidence method, ideas about light quanta in a paper on light quanta and interference.

In the same year he began to study the transformation of light elements by bombardment with alpha rays. The resulting fission products had, until then, been seen by the eye only as scintillations, but Bothe, in collaboration with Fränz, made it possible to count them by means of their needle counter.

In 1929, in collaboration with W. Kolhörster, Bothe introduced a new method for the study of cosmic and ultraviolet rays by passing them through suitably arranged Geiger counters, and by this method demonstrated the presence of penetrating charged particles in the rays, and defined the paths of individual rays.

For his discovery of the method of coincidence and the discoveries subsequently made by it, which laid the foundations of nuclear spectroscopy, Bothe was awarded, jointly with Max Born, the Nobel Prize for Physics for 1952.

In 1930 Bothe, in collaboration with H. Becker, bombarded beryllium of mass 9 (and also boron and lithium) with alpha rays derived from polonium, and obtained a new form of radiation that was even more penetrating than the hardest gamma rays derived from radium, and this led to the discovery of the neutron, made by Sir James Chadwick in 1932.

At Heidelberg, Bothe was able, after much difficulty, to obtain the money necessary for building a cyclotron. He worked, during the 1939–1945 war,

on the diffusion theory of neutrons and on measurements related to these.

In June 1940 he published his *Atlas of Cloud-Chamber Figures*.

He was a member of the Academies of Sciences of Heidelberg and Göttingen, and a Corresponding Member of the Saxon Academy of Sciences, Leipzig. He was awarded the Max Planck Medal and the Grand Cross of the Order for Federal Services. In 1952, he was made a Knight of the Order of Merit for Science and the Arts.

Bothe's remarkable gifts were not restricted to physics. He had an astonishing gift of concentration and his habit of carefully making the best use of his time enabled him to work at great speed. In the laboratory he was often a difficult and strict master, at his best in discussions in small classes there, but in the evenings at home he was, with his Russian wife, very hospitable and all the difficulties of the day were then forgotten.

To his hobbies and recreations he gave the same concentration and intensity of effort that he gave to his scientific work. Chief among them were music and painting. He went to many musical concerts and himself played the piano, being especially fond of Bach and Beethoven. During his holidays he visited the mountains and did many paintings in oil and water colour. In these his style was his own. He admired the French impressionists and was eager and vigorous in his discussions of the merits and demerits of various artists.

Bothe married Barbara Below of Moscow. Her death preceded his by some years. They had two children.

Physics 1955

WILLIS EUGENE LAMB

«for his discoveries concerning the fine structure of the hydrogen spectrum»

POLYKARP KUSCH

«for his precision determination of the magnetic moment of the electron»

Physics 1955

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

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

The Nobel Prize winners in Physics of this year were both employed shortly before the war at the Physics Laboratory of Columbia University in New York. Lamb was at first engaged in theoretical research and published several important investigations in this field. Kusch was soon one of the most active collaborators of Professor Rabi, when the latter worked out his resonance method, for which he was awarded the Nobel Prize in Physics of 1944. In this method the spectra of the atoms are studied by radio waves and the details of the spectra can thereby be investigated much more accurately than before. Kusch and Lamb participated during the war in the extensive work on radar technique which was then being performed. Because of the great progress in this field the resonance method could be much improved. It was used by Kusch and, in an essentially modified form, by Lamb when in 1947 they made their great discoveries as leaders of two separate research groups. Not only were the discoveries made independently of each other in the same laboratory and in the same year, but it was also soon found that the explanation of both phenomena is the same, namely the interaction of the electrons and the electromagnetic radiation.

Lamb's discovery refers to the hydrogen atom, where one single electron moves around the nucleus in one of a series of orbits, each having a definite energy. These energy levels exhibit a fine structure which means that they are arranged in groups of neighbouring levels, the groups being widely separated. An explanation of the fine structure which for a long time was thought to be correct, was given in 1928 by the English physicist Dirac, when he proposed a theory of the electron based on the requirements of the theory of relativity as well as the quantum theory.

Using optical methods many attempts were made during the next decade to check the Dirac theory of the fine structure but no definite results were obtained. Some investigations made it probable, however, that the theory was not entirely correct and the guess was made that there occur deviations which are similar to those later found by Lamb.

Lamb was aware of the great importance that a careful check of the Dirac theory would have. He began his experimental investigations of the fine structure shortly after the war. His technique was based on Rabi's resonance method which had to be much modified, however. Lamb planned his difficult experiment guided by a thorough theoretical analysis of the experimental arrangements.

In 1947 his experiments were successful. He found that two fine structure levels in the next lowest group which should coincide according to the Dirac theory are in reality shifted relative to each other by a certain amount which is now called the Lamb shift. He succeeded in measuring this shift with great accuracy and later made similar measurements on heavy hydrogen.

The discovery of Kusch refers directly to an important property of the electron, namely its magnetic moment. It had been known since long that the electron is a small magnet. The strength of this magnet is measured by its moment. The magnitude of the moment should be uniquely determined by the electron theory of Dirac, mentioned before.

At the beginning of 1947 Rabi together with several collaborators found that a property of the lowest hydrogen level (namely its so-called hyperfine structure) does not entirely conform with theory. It was suggested by the American physicist Breit that the reason for this could be that the magnetic moment of the electron is somewhat different from the value assumed until then which is called a Bohr magneton.

Starting from this idea Kusch made a series of very careful investigations and found in 1947 that the magnetic moment of the electron is larger than the Bohr magneton by about one part in a thousand.

The effects discovered by Lamb and Kusch are exceedingly small. They were revealed only with the help of a very refined technique. As has happened before it was now found that the discovery of minute deviations from existing theories can be of far-reaching importance. The discoveries of Lamb and Kusch resulted in a reshaping of the theory of the interaction of the electrons and the electromagnetic radiation, the so-called quantum electrodynamics.

Lamb reported on his results at a physics meeting which was held in the neighbourhood of New York in the early summer of 1947. Many prominent theoretical physicists were present and among them Professor Kramers from Holland, who died a few years ago. During the discussion it was made probable that the Lamb shift could be explained using certain general ideas of Kramers, the purpose of which was to improve the theories just mentioned.

A rough estimate which agreed rather well with Lamb's measurements was soon made and somewhat later Lamb himself and many others carried out more accurate calculations. It was also found by Professor Schwinger at Harvard University that the anomaly in the magnetic moment of the electron found by Kusch could be similarly explained. In both cases the measurements as well as the calculations have since been considerably improved and agree now very well.

Professor Willis Lamb. Professor Polykarp Kusch. Your discoveries which the Royal Swedish Academy of Sciences wishes to recognize on this occasion have been made by applying radiofrequency spectroscopy of the highest achievable precision to the study of the properties of the electron. Your work is marked not only by the beauty of your experiments but equally by the profound significance of your results. It does not often happen that experimental discoveries exert an influence on physics as strong and invigorating as did your work. Your discoveries led to a re-evaluation and a re-shaping of the theory of the interaction of electrons and electromagnetic radiation, thus initiating a development of utmost importance to many of the basic concepts of physics, a development the end of which is not yet in sight.

I now ask you to receive your Nobel Prize from the hands of His Majesty the King.

WILLIS E. LAMB, JR.

Fine structure of the hydrogen atom

Nobel Lecture, December 12, 1955

When the Nobel Prizes were first awarded in 1901, physicists knew something of just two objects which are now called « elementary particles »: the electron and the proton. A deluge of other « elementary » particles appeared ater 1930; neutron, neutrino, μ meson, π meson, heavier mesons, and various hyperons. I have heard it said that « the finder of a new elementary particle used to be rewarded by a Nobel Prize, but such a discovery now ought to be punished by a $10,000 fine ».

In order to determine the properties of elementary particles experimentally it is necessary to subject them to external forces or to allow them to interact with each other. The hydrogen atom which is the union of the first known elementary particles: electron and proton, has been studied for many years and its spectrum has taught us much about the electron.

In 1885, Balmer found that the wavelengths of fourteen lines of the hydrogen spectrum were given by a simple equation. In 1887, Michelson and Morley discovered a fine structure of some of these lines. The quantum theory was founded by Planck in 1900, and in 1913 Bohr gave rules of quantization which permitted a derivation of Balmer's formula. Sommerfeld showed in 1916 that the fine structure of Bohr's energy levels was caused by relativistic corrections. In 1924, De Broglie attributed wave properties to the electron and soon a quantum mechanics of the hydrogen atom emerged from the hands of Heisenberg, Born, and Schroedinger. Spin and magnetic moment of the electron were suggested by Uhlenbeck and Goudsmit in 1925, and their dynamical equations were worked out by Thomas a year later. In 1928, Dirac discovered an equation which described an electron with wave properties, charge, spin, magnetic moment and a mass depending on velocity as required by relativity theory. The energy levels of hydrogen were given by Dirac's theory with high precision.

Of special interest to us are his predictions, as shown in Fig. 1, of the $n = 2$ group of energy levels which are 10.2 electron volts above the $n = 1$ ground state. The fine structure splitting $2^2P_{3/2} - 2^2P_{1/2}$, which according to Dirac's theory arises from spin-orbit interaction, agrees exactly with the sep-

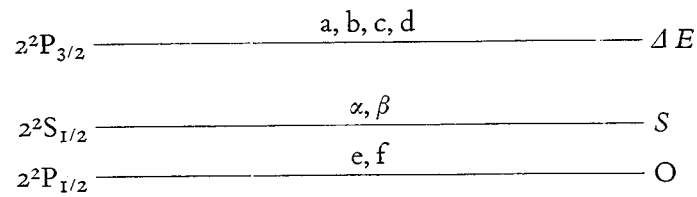


Fig. 1. Fine structure of $n = 2$ levels of hydrogen. According to the Dirac theory the $2^2S_{1/2}$ and $2^2P_{1/2}$ levels coincide exactly. The letters a, b, c, d, e, f, α and β denote the magnetic sub-levels which are split in a magnetic field.

aration of the two levels of Sommerfeld's 1916 theory. The exact coincidence in energy of the $2^2S_{1/2}$ and $2^2P_{1/2}$ states is a consequence of the assumed Coulomb law of attraction between electron and proton. Any departure from this law would cause a separation of these levels.

Many spectroscopic studies of the hydrogen fine structure were made to test the Dirac theory, but by 1940 had failed to establish clearly a decision, although there was evidence strongly supporting the theory. (We now know that the work of Houston[1] and Williams[2] indicated a discrepancy which should have been taken seriously.)

For the subsequent developments, some chapters from my own peculiar history may be of interest. After undergraduate training as a chemist, I studied theoretical physics under Professor J. R. Oppenheimer at the University of California from 1934 to 1938. My thesis[3] dealt with field theories of nucleons which predicted a very small discrepancy from Coulomb's law about a proton. At Columbia University after 1938, I came into close relation with Professor I. I. Rabi and members of the molecular beam laboratory. My attention was drawn briefly to metastable atoms[4] in connection with a proposed atomic beam experiment. During the war, at the Columbia Radiation Laboratory, I received some first-hand acquaintance with microwave radar and vacuum-tube construction techniques. One of the wartime projects in the Laboratory was the determination[5] of the absorption coefficient of centimeter waves in atmospheric water vapor, and my interest was started in what was to become the very active postwar field of microwave spectroscopy.

In teaching a summer session class in atomic physics in 1945 using a textbook[6] by Herzberg, I found references to some attempts[7] made in 1932–1935 to detect absorption of short-wavelength radio waves in a gas discharge of atomic hydrogen. At first it seemed to me that these experiments had failed because of inadequate microwave techniques. I thought of repeating them

with the greatly improved facilities developed during the war. On further study, however, I realized that atoms in a gas discharge are strongly perturbed and that all the $n=2$ states would have nearly equal populations. Under these conditions there would be no appreciable absorption of radio waves caused by transitions between the three states of interest.

It took almost a full year before a workable scheme was clear in my mind. I considered making use of the possible metastability of the $2^2S_{1/2}$ state of hydrogen. In simplest terms this state should be long-lived against radiative transition to the ground state because a spherically symmetrical charge and current distribution cannot radiate according to classical electromagnetic theory. Nevertheless, quite a few papers between 1924 and 1933 were devoted to discussion and experiment to determine whether or not the $2^2S_{1/2}$ state was actually metastable. In 1933, Bethe[8] showed it to be metastable only for the atom in a region sufficiently free of electric fields. However, it was by no means clear that the excitation of the $2^2S_{1/2}$ state could be carried out without destroying the metastability of the state. It was still necessary to detect any interaction of microwaves with the excited atoms, and as already mentioned, a direct absorption method applied to a discharge seemed out of the question. I decided to try to form a beam of metastable hydrogen atoms. If a radio-induced transition to $2^2P_{1/2}$ or $2^2P_{3/2}$ did take place, the atom would lose its metastability and go in about 10^{-9} sec to the ground state with emission of radiation. The beam of metastable atoms would thereby be diminished.

It had been showed in 1924 by Webb[9] at Columbia University that metastable mercury atoms could liberate electrons from metals, but no one had either produced or detected the strange metastable form of hydrogen, and it was not certain that a beam of it could be detected through its ejection of electrons.

On the basis of these rather doubtful prospects, I persuaded Retherford[10] to attempt the experiment represented schematically in Fig. 2. After several failures, a successful apparatus was made as shown in Fig. 3. Molecular hy-

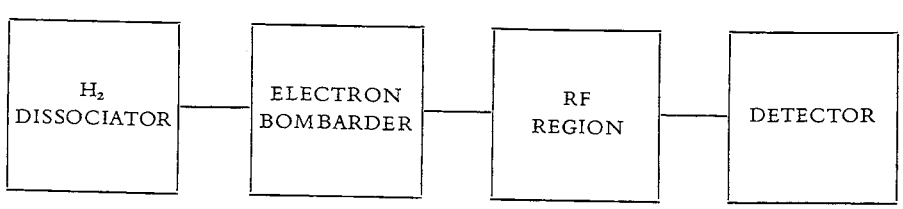


Fig. 2. Modified schematic block diagram of apparatus.

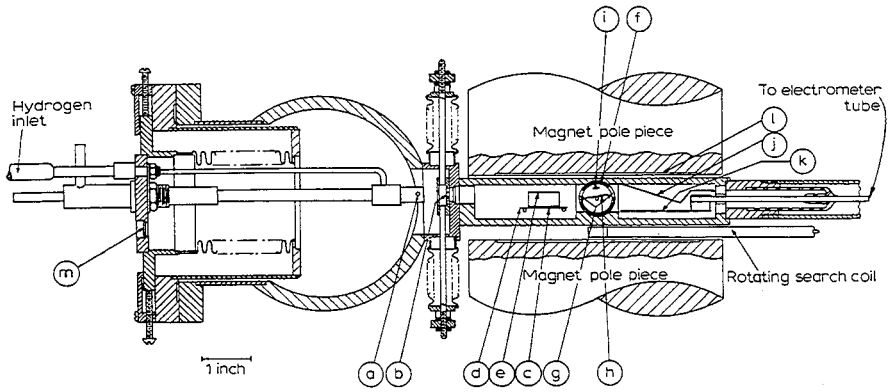


Fig. 3. Cross section of second apparatus: (*a*) tungsten oven of hydrogen dissociator, (*b*) movable slits, (*c*) electron bombarder cathode, (*d*) grid, (*e*) anode, (*f*) transmission line, (*g*) slots for passage of metastable atoms through interaction space, (*h*) plate attached to center conductor of r-f transmission line, (*i*) d.c. quenching electrode, (*j*) target for metastable atoms, (*k*) collector for electrons ejected from target, (*l*) pole face of magnet, (*m*) window for observation of tungsten oven temperature.

drogen is dissociated in a tungsten oven, and a stream of hydrogen atoms emerges from a slit to be bombarded by electrons which bring about one atom in a hundred million atoms to the metastable state. After a small recoil deflection, the excited atoms move on to a metal surface from which they can eject electrons and so be detected. Between bombarder and detector, the atoms are exposed to radio waves.

For several good reasons the whole process is carried out in a magnetic field. The fine structure energy levels are subject to Zeeman splitting as shown in Fig. 4, and the frequencies of the possible transitions depend on magnetic field as shown in Fig. 5. As the magnetic field is varied with radio-frequency and amplitude held fixed $2S \rightarrow 2P \rightarrow 1S$ transitions occur and a certain fraction of metastable atoms in the beam are quenched and lose their excitation energy. Since atoms in the ground state are unable to eject electrons from the detector, the electrometer current decreases. A typical resonance curve for deuterium is shown in Fig. 6 and for hydrogen in Fig. 7. The widths of the resonance peaks are partly due to natural radiative decay of the 2P state, and partly to hyperfine structure which is just resolved for one of the hydrogen resonances.

By careful experimental study and theoretical analysis of such resonance curves, it is possible to determine the energy separations $2^2S_{1/2} - 2^2P_{1/2}$ and $2^2P_{3/2} - 2^2P_{1/2}$ in zero magnetic field. The accurate measurements[11] by Day-

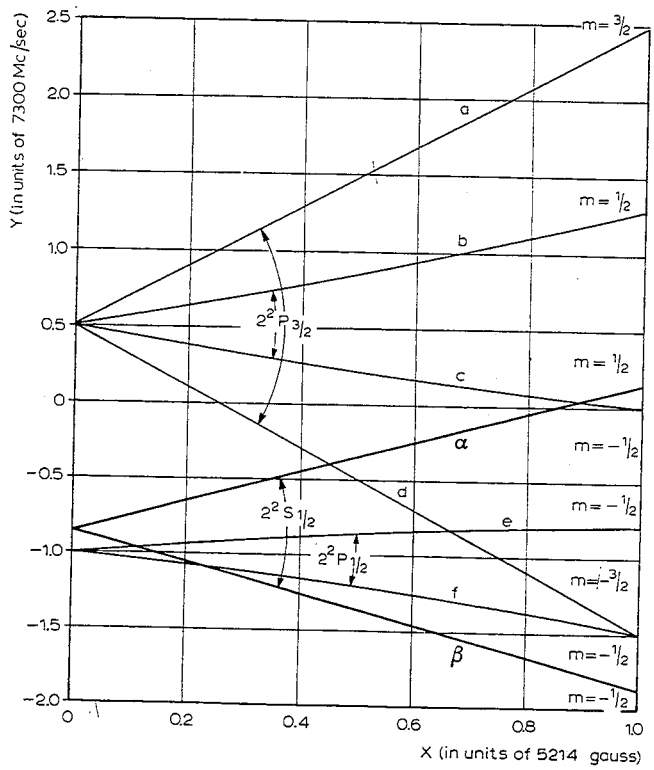


Fig. 4. Zeeman energy levels with the $2^2S_{1/2}$ pattern raised by 1000 Mc/sec.

hoff, Triebwasser, and Lamb gave 1059.0 and 9912.6 Mc/sec for these separations in deuterium with an accuracy of about 0.1 Mc/sec.

Studies of other fine structure states have been made by microwave methods. For the $n = 2$ state of singly ionized helium, the separations are 13 to 16 times larger than for hydrogen. In the work of Lamb and Skinner[12], no beam of metastable ions was formed, but rather the ultraviolet radiation emitted in the decay of 2P — 1S was used to detect the transitions 2P — 2S induced by microwaves. A similar method was used in the later experiments of Deutsch[13] on positronium in which a transition from the triplet to singlet state changes the rate of emission of annihilation radiation. Recently the fine structure of the $n = 3$ state of hydrogen was studied by Lamb and Sanders[14] using analogous microwave methods, and the fine structure of the 3^3P state of helium has been determined[15] by Lamb, Maiman, and Wieder.

Let me now tell briefly about the explanation which has been given of the departures from the expected fine structure pattern for hydrogen. To come

at once to the heart of the matter: the electron does not behave like a point charge as implied in Dirac's equation. If time permitted, I would trace the long history of attempts to make a theory of the internal structure of electrons. This was begun by J. J. Thomson with his calculation of electromagnetic mass, and developed by Lorentz using classical electromagnetic theory.

According to relativity theory, energy and mass are proportional to one another. In order to obtain a finite mass of the electron on a purely electromagnetic basis, it was necessary to assign an extended structure to the electron. Attempts to form a satisfactory relativistic theory of an extended charged particle failed.

After quantum mechanics had been applied to matter it was used by Dirac and by Heisenberg and Pauli to attack the problem of the electromagnetic field and its interaction with charged particles. This theory had many successes. A good account was given of the phenomena of emission, absorption

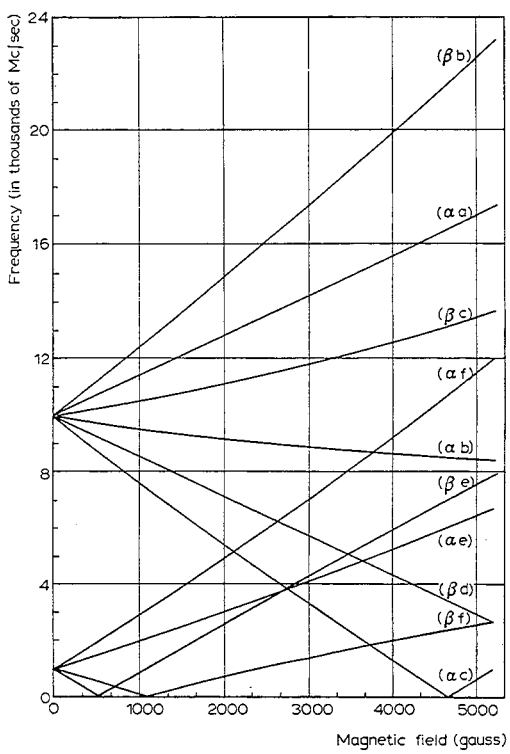


Fig. 5. Expected resonance frequencies as functions of magnetic field with the $2^2S_{1/2}$ state raised by 1000 Mc/sec.

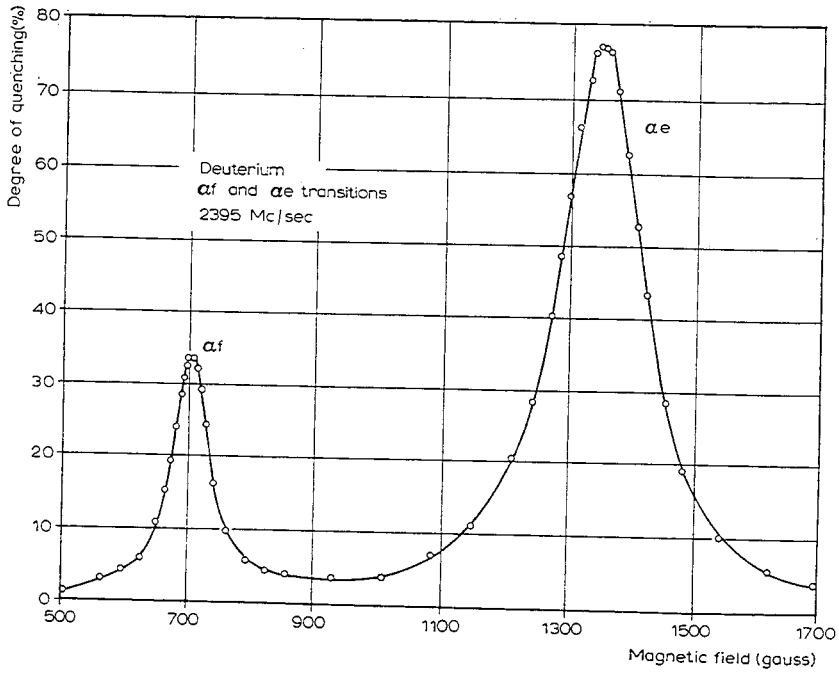


Fig. 6. Observed resonance curves for deuterium.

and scattering of radiation by matter, including the spontaneous radiation postulated by Einstein in 1917. From the quantum electrodynamical point of view, spontaneous emission could be regarded as emission stimulated by the quantum zero-point fluctuations in the electromagnetic fields in the vacuum or lowest energy state.

When, however, the energy of interaction of an electron with the quantized electromagnetic field was calculated by Oppenheimer[16] in 1930, a meaningless result was obtained. According to this theory, not only was the energy infinite but the frequencies of all spectral lines were displaced by an inifinite amount from the values predicted without consideration of radiative interactions. It was pointed out by Weisskopf[17] in 1934 that this ultraviolet catastrophe could be ascribed to the above-mentioned zero-point fluctuations in the vacuum electromagnetic field. These fluctuations were needed for the correct spontaneous emission to occur and yet led to violent disagreement with observations in other respects. The theory of electromagnetic energy remained in this highly unsatisfactory state until May 1947 when the fine structure deviations were definitely established experimentally.

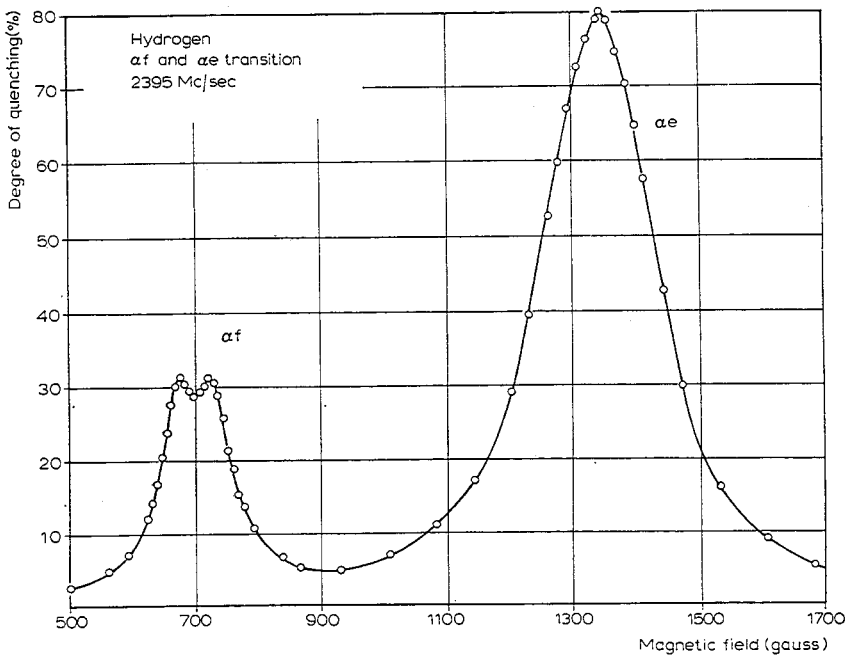


Fig. 7. Observed resonance curves for hydrogen.

A month later, Bethe[18] found that quantum electrodynamics had really hidden behind its divergencies a physical content which was in very close agreement with the microwave observations. The crucial step due to Bethe was a successful application of the very old idea of renormalization of mass. The quantum theory of radiation predicted that a free electron should have an infinite mass. Some explanation must therefore be found for the observed fact that the mass was finite. Bethe recognized that this was outside the scope of the theory at that time and simply ignored the electromagnetic mass difficulty. For an electron bound in a hydrogen atom, an infinite energy also occurs, but this was merely a manifestation of the infinite electromagnetic mass which should be eliminated in some future theory. With the mass terms properly subtracted a finite remainder is obtained which would be zero for a free electron. In the case of a bound electron, the force field modifies the effect of the electromagnetic field and a finite displacement of energy levels results which is in quite good accord with observation.

A qualitative picture of the level shift was given by Welton[19] in 1948. The fluctuating zero-point electric field of the quantized vacuum acts on an elec-

tron bound in a hydrogen atom. As a result, the electron is caused to move about its unperturbed position in a rapid and highly erratic manner. The point electron effectively becomes a sphere of a radius almost 10^{-12} cm. Such an electron in a hydrogen atom is not so strongly attracted to the nucleus at short distances as would be a point electron. States of zero orbital angular momentum like $2^2S_{1/2}$ are therefore raised in energy relative to other states like 2^2P in which the electron has a smaller probability of being found near the nucleus.

In 1949, a relativistic generalization of Bethe's calculation was given by Kroll and Lamb[20], which made his results definite. They confirmed additional small contributions of 27 Mc/sec arising from polarization of the vacuum as calculated in 1935 by Uehling[21] on the basis of Dirac's theory of the positron, and of 68 Mc/sec from the anomalous magnetic moment of the electron as suggested by Breit[22] in 1947. Other small corrections have been calculated by various authors, of which the largest was about 7 Mc/sec made by Baranger[23] who took the binding of the electron more exactly into account. At the present time, there is an unexplained residual discrepancy of 0.5 Mc/sec.

It is very important that this problem should receive further experimental and theoretical attention. When an accuracy of comparison of 0.1 Mc/sec has been reached, it will mean that the energy separations of the 2S and 2P states of hydrogen agree with theory to a precision of a few parts in 10^9 of their binding energy or that the exponent in Coulomb's law of force is two with a comparable accuracy. Another way of putting it is to say that the anomalous magnetic moment of the electron would be determined with an accuracy of one part in 680, which would provide a useful check on Kusch's more directly measured result[24]. Finally, I might mention that the fine structure doublet separation now provides the most accurate and direct determination of the famous dimensionless number called the fine structure constant, whose numerical value of about 1/137 it will be the task of some future theory to explain.

1. W. V. Houston, *Phys. Rev.*, 51 (1937) 446.
2. R. C. Williams, *Phys. Rev.*, 54 (1938) 558.
3. W. E. Lamb, Jr. and L. I. Schiff, *Phys. Rev.*, 53 (1938) 651.
4. A. Cobas and W. E. Lamb, Jr., *Phys. Rev.*, 65 (1944) 327.
5. G. E. Becker and S. H. Autler, *Phys. Rev.*, 70 (1946) 300; W. E. Lamb, Jr., *Phys. Rev.*, 70 (1946) 308.
6. G. Herzberg, *Atomic Spectra and Atomic Structure*, Prentice-Hall, Inc., New York, 1937.
7. O. Betz, *Ann. Physik*, 15 (1932) 321 and T. Haase, *Ann. Physik*, 23 (1935) 675.
8. H. A. Bethe, *Handbuch der Physik*, 2nd ed., Vol. 24/1, 1933, pp. 452–462.
9. H. W. Webb, *Phys. Rev.*, 24 (1924) 113.
10. W. E. Lamb, Jr. and R. C. Retherford, *Phys Rev.*, 72 (1947) 241; 79 (1950) 549; 81 (1951) 222. W. E. Lamb, Jr., *Phys. Rev.*, 85 (1952) 259. W. E. Lamb, Jr. and R. C. Retherford, *Phys. Rev.*, 86 (1952) 1014.
11. S. Triebwasser, E. S. Dayhoff, and W. E. Lamb, Jr., *Phys. Rev.*, 89 (1953) 98. E. S. Dayhoff, S. Triebwasser and W. E. Lamb, Jr., *Phys. Rev.*, 89 (1953) 106.
12. W. E. Lamb, Jr. and Miriam Skinner, *Phys. Rev.*, 78 (1950) 539.
13. M. Deutsch and E. Dulit, *Phys. Rev.*, 84 (1951) 601 and M. Deutsch and S. C. Brown, *Phys. Rev.*, 85 (1952) 1047.
14. W. E. Lamb, Jr. and T. M. Sanders, Jr., *Phys. Rev.*, 103 (1956) 313. Also, W. E. Lamb, Jr., *Proc. Am. Acad. Arts Sci.*, 82 (1953) 344.
15. T. H. Maiman and W. E. Lamb, Jr., *Phys. Rev.*, 98 (1955) 1194(A) and other articles in preparation. (*Phys. Rev.*, 105 (1957) 573.)
16. J. R. Oppenheimer, *Phys. Rev.*, 35 (1930) 461.
17. V. F. Weisskopf, *Z. Physik*, 90 (1934) 817.
18. H. A. Bethe, *Phys. Rev.*, 72 (1947) 339.
19. T. A. Welton, *Phys. Rev.*, 74 (1948) 1157.
20. N. M. Kroll and W. E. Lamb, Jr., *Phys. Rev.*, 75 (1949) 388.
21. E. A. Uehling, *Phys. Rev.*, 48 (1935) 55.
22. G. Breit, *Phys. Rev.*, 72 (1947) 984.
23. M. Baranger, *Phys. Rev.*, 84 (1951) 866.
24. P. Kusch, *Nobel Lecture*, this volume, p. 298.

Biography

Willis Eugene Lamb, Jr. was born on July 12, 1913 in Los Angeles, California. His father Willis Eugene Lamb, born in Minnesota, was by profession a telephone engineer and his mother Marie Helen Metcalf came from Nebraska.

Except for three years schooling in Oakland, Calif., he was educated in the public schools of Los Angeles, Calif. In 1930 he entered the University of California at Berkeley and received a B.S. (Chemistry) in 1934. His graduate work in theoretical physics at the same university led to the Ph.D. degree in 1938. His thesis research on the electromagnetic properties of nuclear systems was directed by Professor J. R. Oppenheimer.

He went to Columbia University as Instructor in Physics in 1938, became an Associate (1943), Assistant Professor (1945), Associate Professor (1947) and Professor in 1948. From 1943 to 1951, he was associated also with the Columbia Radiation Laboratory where the research described in the Nobel Lecture was done. In 1951 he went to Stanford University in California as Professor of Physics. During 1953–1954 he was Morris Loeb Lecturer at Harvard University. From 1956 to 1962 he was a Fellow of New College and Wykeham Professor of Physics at the University of Oxford, England. In 1962 he became Henry Ford II Professor of Physics at Yale University, New Haven, Conn.

His research has been on the following subjects: theory of the interactions of neutrons and matter, field theories of nuclear structure, theories of beta decay, range of fission fragments, fluctuations in cosmic ray showers, pair production, order-disorder problems, ejection of electrons by metastable atoms, quadrupole interactions in molecules, diamagnetic corrections for nuclear resonance experiments; theory and design of magneton oscillators, theory of a microwave spectroscope, study of the fine structure of hydrogen, deuterium and helium; theory of electrodynamic energy level displacements.

In 1953 he received the Rumford Premium of the American Academy of Arts and Sciences. The University of Pennsylvania conferred an honorary

degree of D.Sc. upon him in 1954. He received the Research Corporation Award in 1955. He is a member of the National Academy of Sciences, and a Fellow of the American Physical Society.

In 1939 he married Ursula Schaefer, a student from Germany.

POLYKARP KUSCH

The magnetic moment of the electron

Nobel Lecture, December 12, 1955

I must tell you, and with considerable regret, that I am not a theoretical physicist. A penetrating analysis of the part that the discovery and measurement of the anomalous magnetic moment of the electron has played in the development of certain aspects of contemporary theoretical physics must be left to the group of men who have in recent years devised the theoretical structure of quantum electrodynamics. My role has been that of an experimental physicist who, by observation and measurement of the properties and operation of the physical world, supplies the data that may lead to the formulation of conceptual structures. The consistency of the consequences of a conceptual structure with the data of physical experiment determines the validity of that structure as a description of the physical universe. Our early predecessors observed Nature as she displayed herself to them. As knowledge of the world increased, however, it was not sufficient to observe only the most apparent aspects of Nature to discover her more subtle properties; rather, it was necessary to interrogate Nature and often to compel Nature, by various devices, to yield an answer as to her functioning. It is precisely the role of the experimental physicist to arrange devices and procedures that will compel Nature to make a quantitative statement of her properties and behavior. It is in this spirit that I propose to discuss my participation in a sequence of earlier experiments that made possible the precision determination of the magnetic moment of the electron. I will then discuss the experiments themselves which have yielded our present knowledge of the magnetic properties of the electron.

Research with atomic and molecular beams has had a long and fruitful record in the history of the growth of our present knowledge of matter. The experiments that I shall discuss are some in which the method of atomic and molecular beams is used essentially as a spectroscopic device for the observation of spectral lines in the range of frequencies within which power may be generated by electronic means. The general principles of radiofrequency spectroscopy by the method of molecular beams were first described by Rabi[1] and a group of his co-workers of which I was fortunate to be a mem-

ber. It is here sufficient to say that a transition between energy levels may be observed through the circumstance that the magnetic moment of an atom or molecule may be changed in a transition. The method is characterized by a very high potential resolution, and in many observations of the frequency of a line, an accuracy of better than one part in a million has been achieved. It is of particular value as a tool in the investigation of the details of interactions within atoms and molecules because small interactions appear as first-order effects rather than as small superpositions on the relatively enormous energies that characterize optical spectra.

The fact that the electron has a spin of one half and a magnetic moment at least approximately equal to one Bohr magneton has long been recognized. Uhlenbeck and Goudsmit[2] first postulated these properties of the electron to explain the fine structure in atomic spectra and what has been called the anomalous Zeeman effect. An enormous body of evidence has given an ever-increasing support to these postulates. The relativistic Dirac theory of the electron assumed a particle that was endowed with the properties of mass and charge. The spin and magnetic moment postulated by Uhlenbeck and Goudsmit were then found to be a consequence of the relativistic invariance of the Dirac equation. Indeed, one of the great triumphs of the Dirac electron theory was the prediction of these postulated electron properties. The spin and moment of the electron were thus removed from the realm of *ad hoc* assumptions, justified by experimental evidence, to the realm of an integral part of quantum theory. The Dirac electron theory did not, however, consider the interaction of the quantized electromagnetic field with the electron.

I shall talk of the measurement of the g value rather than the magnetic moment of the electron. The g value is, as usual, the negative ratio of the magnetic moment in terms of the Bohr magneton μ_0 and the angular momentum in units of $h/2\pi$. Since, in all cases here under discussion, the angular momentum of the system is known, the moment can immediately be obtained from the g value. The most elementary of the g values, g_L, is that associated with the orbital motion of the electron. Its value is 1 within small and calculable corrections. The electron also has a magnetic moment by virtue of its angular momentum about a spin axis. The g value associated with the spin, g_S, is the quantity here under investigation; a value of 2 was obtained for it in the Dirac electron theory. Now the electrons in an atom have both spin and orbital angular momentum. To the total electronic angular momentum J, we assign the g value g_J. The atom contains a nucleus that may have a nuclear angular momentum and hence a nuclear magnetic moment.

The nuclear g value, g_N, is designated as g_P in the special case when the nucleus is a proton. To the total angular momentum of the atom we assign the g value g_F.

The earliest measurements by the molecular beams magnetic resonance method were undertaken by a group[1] of which I was a member working in Professor Rabi's laboratory and under his direction. The measurements consisted of the determination of nuclear g values by the observation in a molecule of the nuclear resonance frequency in a classically determined magnetic field. Even in the great national laboratories dedicated to the maintenance of physical standards, a precision of only about one part in forty thousand has been achieved[3] in the measurement of a field. In a well-equipped laboratory that, however, lacks the equipment and tradition of meticulous intercomparison of electrical standards, a precision of perhaps one quarter percent may be achieved in the determination of the magnitude of a field. While ratios of nuclear g values may be found without an explicit knowledge of the field, the accuracy of the determination of a nuclear moment in terms of the Bohr or nuclear magneton is limited by the uncertainty in the measurement of a field as well as by the uncertainty in a prior measurement of the Bohr or nuclear magneton. Thus, the desirability of a direct measurement of a nuclear g value in terms of the Bohr magneton is apparent.

The molecular beam magnetic resonance method was originally applied to the determination of the nuclear g values in molecules that did not have a net electronic orbital or spin angular momentum. It is, however, possible to apply the same experimental techniques to an investigation of the hyperfine structure of atoms. If we observe transitions between the various F levels at zero or very low magnetic field, the hyperfine structure separation may readily be found. At a higher magnetic field the observation of the frequency of transition between magnetic sub-levels yields again the zero-field h.f.s. (hyperfine structure) splitting, the quantity $g_J\mu_0 H/h$ and the quantity g_N/g_J, though the latter quantity can be found with only limited precision. The group of which I was a member at Columbia made[4] the first such studies on the commonly occurring isotopes of the alkali atoms and determined the magnetic hyperfine structure interaction constants of the alkali atoms. Extensive subsequent work in the observation of atomic h.f.s. has, of course, been done in many laboratories with results of great interest in the study of higher-order moments in nuclei and of the properties of radioactive nuclei. The alkali atoms were particularly adaptable to the original experimental work, first because the beams of the atoms may readily be produced and de-

tected and secondly, because they occur in $^2S_{1/2}$ states, almost wholly free of perturbation by other states.

The possibility of measuring the moment of a nucleus in terms of the Bohr magneton is a consequence of the possibility of observing both nuclear resonance in molecules with a frequency $g_N\mu_0 H/h$ and transitions among the magnetic components of h.f.s. levels for which the dependence of frequency on field is of the order of $g_J\mu_0 H/h$. Millman and I[5] then addressed ourselves to the problem of measuring the moment of the proton in terms of the electron spin moment. The experimental problems were considerable and arose from three factors. The first of these was related to the fact that the effective moment of a molecule of zero electronic angular momentum is of the order of a nuclear magneton while that of an atom is of the order of a Bohr magneton. Deflecting fields that allow the observation of a change in trajectory of a molecule in which a hydrogen nucleus has undergone a transition will deflect an atom through unmanageably large excursions if the field is arbitrary. However, all atoms in which magnetic h.f.s. occurs and in which the spin of the nucleus is greater than 1/2 have, in certain states, zero magnetic moments at definite values of the magnetic field which may be very high. Atoms in such states may thus traverse a carefully adjusted inhomogeneous field without catastrophic deflections, and a transition from such a state may at once be detected since the terminal state is generally characterized by a large magnetic moment. Thus, it is possible to choose the deflecting field in such a way that a change in the spin orientation of a nucleus in a molecule and a transition among the magnetic levels of the h.f.s. may both be detected. The second of our experimental problems was related to the production of a beam of molecules that contained hydrogen and simultaneously an alkali atom, requisite for the detection of beams by techniques then available to us. Since atomic lines were to be observed at the same field as nuclear resonance lines, the simultaneous production of a beam of alkali atoms was necessary. We used beams of sodium and potassium hydroxides evaporated from silver ovens and noted that at the temperatures required to generate a beam of the hydroxide, the reaction between an alkali halide and metallic calcium proceeded at such a rate that a convenient beam of atoms appeared. The third experimental problem was associated with the need of applying to the same circuit, in succession, two frequencies that differed by as much as a factor of seventy, and with the pre-war difficulty of generating r.f. (radio-frequency) power at high frequencies.

Extensive intercomparison of the frequencies of the resonance line of the

protons and of lines in the h.f.s. spectra of sodium, rubidium and caesium for each of which a prior determination of the interaction constants had been made, led to a determination of the ratio of the proton moment and the spin magnetic moment of the electron in the calibrating atoms which we, of course, assumed to be the Bohr magneton. The magnetic moment of the proton in terms of the nuclear magneton found on the basis of the assumption that the spin moment of the electron was indeed the Bohr magneton differed from the moment as determined from the measurement of a frequency in a classically determined field by about one tenth percent. When, at a much later date, it was found that the spin magnetic moment of the electron deviates from the Bohr magneton by the order of one tenth percent, the direction of the deviation in the older experiment was examined. It is perhaps a good commentary on the hazards of experimental physics that no significant effect had escaped us but that the error in the mutual inductance which we had used in the calibration of the magnetic field was of the order of two tenths rather than one tenth percent.

After the war, Nafe and Nelson[6], working with Rabi, made the first of the measurements of the hyperfine structure splitting of hydrogen in the ground state. Now the h.f.s. of hydrogen may be calculated explicitly in terms of the magnetic moment of the proton, the spin magnetic moment of the electron and the electronic wave function at the nucleus. However, a discrepancy of about one quarter percent was noted between the observed and predicted magnitude of the h.f.s. splitting when the value of the proton moment found by Millman and me was used. The assumption that the spin moment of the electron is a Bohr magneton enters into the calculation twice, first as an intrinsic property of the electron and second in the calculation of the proton moment from the observed ratio of the proton moment and the spin moment of the electron. The discrepancy led Breit[7] to suggest that the electron may possess an intrinsic magnetic moment greater than μ_0 by the order of $\alpha\mu_0$, where α is the usual fine structure constant.

The question of the existence of an anomalous magnetic moment was then investigated in detail by Foley and me[8]. In this inquiry, as in all others conducted in the atomic and molecular beams laboratory at Columbia University, we profited by Rabi's advice. The procedure that we employed capitalizes on the fact that the g_J value associated with a state is a linear combination of the electronic orbital and spin g values, g_L and g_S, and that this combination is different for different states. That is, there is a contribution to the total electronic magnetic moment of an atom both from the orbital mo-

tion of the electrons and from the spin of the electrons, and the contribution from each of these factors is dependent on the state of the atom. Since we considered only atoms with single electrons outside of closed shells, Russell-Saunders coupling is a good approximation and the coefficients that relate the various g values are known. The ratio of the g_J values of two atoms that occur in different spectroscopic states yields g_S/g_L to an accuracy limited by the precision of observation and the precision with which the coefficients relating the various g values are known.

The intercomparison of g values to obtain a value of g_S can only be made if atoms in several different spectroscopic states are available for observation. After our first investigation of the hyperfine structures of the alkali atoms, all in the $^2S_{1/2}$ state, Hardy and Millman[9] studied the h.f.s. of indium in the $^2P_{1/2}$ state. Just after the war Becker and I[10] determined the interaction constants that characterize the h.f.s. of both isotopes of gallium in the $^2P_{1/2}$ state. Gallium atoms in the excited and metastable $^2P_{3/2}$ state also occur in an atomic beam, and it was possible to determine the interaction constants for both isotopes in this state as well. We thus had available for study atoms in three different spectroscopic states.

In principle the determination of the ratio of two g_J values is simple. Suppose we observe transitions for which F is constant and m_F changes by ± 1 for two different atoms or for the same atom in two different states. To the extent to which strictly low field conditions prevail, all lines in a given F state have the same frequency and the ratio of the frequencies of such lines in two different states at a fixed field is simply the ratio of the g_F values. From this ratio, g_S/g_L for the electron may readily be derived with some additional knowledge of the properties of the nucleus in each atom. However, the h.f.s. splitting of atomic states is generally small (from 200 to 20,000 megacycles) and the energies of the levels are far from linearly dependent on magnetic field at usefully high fields. It is, nevertheless, possible to obtain expressions for the energies of all levels in the h.f.s. in terms of the zero field h.f.s. splitting, the ratio g_N/g_J and the quantity $g_J\mu_0 H/h$, or, where such expressions cannot be explicitly found, to determine the energies from the appropriate secular determinants. From the observed frequencies of appropriate lines and with a prior knowledge of the interaction constants which characterize an atom in the state in question it is then possible to determine $g_J\mu_0 H/h$. Measurement of this quantity at the same field for atoms in two different states yields, at once, the important ratio g_S/g_L. The determination is independent of a knowledge of the magnetic field and of any fundamental constants.

It is, perhaps, worthwhile to remark on some experimental details. The field in which the transition frequency was measured was so chosen that all the observed lines had a frequency of the order of 1 megacycle per gauss. To avoid excessive distortion of the lines due to inhomogeneity of the field, a great deal of adjustment of the field was required before the lines approximated in width their theoretical value.

Special arrangements were made to allow the rapid interchange of ovens so that lines of different atoms could be measured in rapid succession. A considerable number of oscillators was required so that several frequencies which differed by large factors could be applied to the r.f. circuits that induced the transitions. While the lines should, in principle, be measured at a fixed if unknown field, the actual measurements were made in a field that varied monotonically throughout a series of observations. The variation of field has the annoying effect of requiring a large body of data to establish the frequencies of two or more lines at a fixed field, but it also aids in avoiding repetitive errors that may occur when a reading of a fixed quantity is repeated.

Three intercomparisons of g_J values were made in these experiments. The results are given in Table 1.

Table 1. Observed ratios of atomic g values and the corresponding values of g_S/g_L

Comparison	*Nominal*	*Observed*	g_S/g_J
$g_J(^2P_{3/2}Ga)/g_J(^2P_{1/2}Ga)$	2	2 (1.00172 ± 0.00006)	2 (1.00114 ± 0.00004)
$g_J(^2S_{1/2}Na)/g_J(^2P_{1/2}Ga)$	3	3 (1.00242 ± 0.00006)	2 (1.00121 ± 0.00003)
$g_J(^2S_{1/2}Na)/g_J(^2P_{1/2}In)$	3	3 (1.00243 ± 0.00010)	2 (1.00121 ± 0.00005)

It is to be noted that the ratio g_S/g_L which has been determined has been found from the ratio of the g_J values on the basis of the assumption that the coupling is Russell-Saunders coupling. Hence the deviation of the ratio g_S/g_L from its nominal value of 2 as determined from any pair of atoms or any pair of states does not constitute clear evidence that the spin moment of the electron is other than one Bohr magneton because of the possibility of occurrence of significant perturbations of the states. Theoretical arguments, however, indicate that such perturbations must be small. On experimental grounds the agreement of the ratio obtained in three different ways from different atoms in different spectroscopic states offers overwhelming evidence that the spin moment of the electron does indeed differ from its nom-

inal value by the indicated amount. The discrepancies between the three values of the ratio may, however, arise from perturbations of the indicated energy levels.

A later intercomparison of the g_J values of the alkali atoms and a comparison of the g_J value of potassium and hydrogen has demonstrated that the g_J values of the three alkali atoms of lowest atomic number are indeed equal to the spin g value of the electron to within one part in forty thousand. A further intercomparison by Mann and me[11] of the g_J values of indium in the $^2P_{1/2}$ and $^2P_{3/2}$ states has given further confirmation to the interpretation of the discrepancy between a measured ratio of g_J values and the nominal value.

The experiments that have been described were performed at a field of about 400 gauss. In a wholly independent series of experiments Taub and I[12] determined the ratio of the g_J value of indium in the $^2P_{1/2}$ state and that of sodium in the $^2S_{1/2}$ state by observations of lines in the h.f.s. spectrum at fields that ranged from 3,300 to 12,000 gauss. The method was to determine the nuclear g value of the proton in an alkali hydroxide in terms of the g_J values of indium and sodium. The result, in so far as it concerns the proton, is of no further interest here in view of the highly refined experiments which have been done in later years that allow the precise and direct determination of the nuclear g value of the proton in terms of both the nuclear and the Bohr magneton. The result is, however, of interest in that it yields again the ratio of the g_J values in two different states on the basis of measurements at fields which differed from those in the earlier experiments by an order of magnitude. We found that:

$$g_S/g_J = 2\,(1 + 0.00119)$$

We may, therefore, conclude on the basis of all evidence that the electron does indeed possess an «intrinsic» or «anomalous» magnetic moment over and above that deduced from the Dirac theory and whose magnitude is very close to 0.119 percent of the Bohr magneton.

Perhaps it is well, at this point, to make a brief statement of the theoretical status of the spin magnetic moment of the electron. Soon after the publication of our first results which gave substance to the assertion that the electron does have an anomalous moment, Schwinger[13] gave a result, based on new procedures in quantum electrodynamics, that

$$g_S/g_L = 2\,(1 + \alpha/2\pi) = 2\,(1.00116)$$

The result is in excellent agreement with experimental measurements of the same quantity. The effect of the increased electron moment arises essentially as a consequence of the quantization of the electromagnetic field which always has a residual zero-point amplitude. While the existence of this field had previously been recognized, it had not been possible to deal with the interaction prior to the formulation of the contemporary quantum electrodynamics. The importance of the observation of the anomalous magnetic moment of the electron is in part in the demonstration that the procedures of quantum electrodynamics are, in fact, satisfactory in formulating a description of Nature.

It is obvious that a more detailed study of the magnetic moment of the electron than that described thus far was desirable. The objective of a more extended investigation lies in the avoidance of theoretical difficulties in the interpretation, to a high precision, of the electronic g values of complex atoms. In the absence of substantial difficulties of interpretation, the very great precision of which spectroscopy by the method of atomic beams is capable may be used to obtain results of sufficient precision to test the validity of the calculations of quantum electrodynamics when made to a higher order than those originally made by Schwinger.

Barring only a measurement of the spin moment of the free electron itself, the best measurement that one may hope to make is on the electron in the hydrogen atom. In this atom in the ground state, the electron has no orbital angular momentum and hence there is no contribution to the electronic magnetic moment from the orbital motion. The entire electronic magnetic moment arises from the spin moment of the electron. Koenig, Prodell and I[14] have determined the ratio of the electronic g value, g_J, of the hydrogen atom and the nuclear g value of the proton by experimental procedures to be described. To a very high order of accuracy, g_J is equal to g_S', the spin g value of the electron bound in the hydrogen atom. The value of g_S' differs from g_S of the free electron through a small relativistic effect of about eighteen parts per million. Corrections due to a mixing of states and relativistic effects are well known and do not limit the accuracy with which the ratio g_S/g_P may be determined at the present time. Gardner and Purcell[15] have measured the ratio $2g_L/g_P$ of the cyclotron frequency of the electron and the precession frequency of the proton in a magnetic field to an accuracy of about one part in eighty thousand. Our result when combined with that of Gardner and Purcell yields g_S/g_L.

As a preliminary procedure, Prodell and I[16] determined the hyperfine

structure separation of hydrogen with high precision. A subsequent investigation by both Wittke and Dicke[17] and by us[18] indicated an excessively optimistic estimate of the uncertainty. However, the value of the zero-field h.f.s. splitting of hydrogen that we used was sufficiently good to contribute no error to the value of g_J/g_P, comparable to other uncertainties.

The apparatus designed for the purpose of these experiments had for the magnet which determined the transition frequencies, one with a much better field homogeneity than that which usually characterizes the magnets used in atomic beams experiments. Ordinarily in an atomic beams apparatus, the magnet that determines the splitting of the levels is internal to the vacuum system. This arrangement permits small magnet gaps to be used and hence the production of large fields with electromagnets of moderate dimensions and power consumption. The use of a small gap, however, leads to a considerable hazard of field inhomogeneity. In the present case the magnet was external to the vacuum envelope, the pole faces were of large diameter to reduce edge effects and the magnet could be carefully shimmed after each change of externally imposed experimental parameters and from day to day to give good homogeneity in the volume within which transitions were observed. The deflecting magnets consisted of current-carrying conductors rather than the iron magnets that have become conventional in atomic beams experiments. This choice was made because of the smaller distortion of the transition field by current-carrying conductors than by massive blocks of iron.

The experiment involved the measurement of the frequency of transition between the levels $m = 0$ and $m = -1$ in the state for which $F = 1$, alternately with the proton resonance frequency in the same magnetic field. The frequency of the first of these lines is of the order of 3,600 megacycles at a field of 1,500 gauss. The frequency of the proton resonance line is about 6.5 megacycles at the same field and is found by the methods of nuclear resonance in the same region of space as that traversed by the beam. An important component of the equipment is a device that can insert a cylindrical sample of water or mineral oil into a region as closely coincident as inherent limitations permit to that in which the atomic line has been observed. Various small corrections relating to the residual inhomogeneity of the field, bulk diamagnetism of the matter in the cylindrical sample that we employed, to the presence of paramagnetic ions when we observed the resonance in water and the differential internal diamagnetic shielding between oil and water must be applied.

We found that

$$g_J/g_P = -658.2171 \pm 0.0006$$

where g_P is the nuclear g value observed in a spherical sample of mineral oil. It is to be noted that this is only an apparent value of g_P since the externally applied field is modified by the internal diamagnetic shielding of the proton by the electrons in the molecules containing the proton. It is, nevertheless, of value to give the result in this form since the ratio $2g_L/g_P$ measured by Gardner and Purcell also refers to a spherical sample of mineral oil.

Application of a small relativistic term yields g_S, the spin g value of the electron in terms of g_P.

$$g_S/g_P = -658.2288 \pm 0.0006$$

The combination of this result with that of Gardner and Purcell

$$2g_L/g_P = -657.475 \pm 0.008$$

yields

$$g_S/g_P = 2\,(1.001146 \pm 0.000012)$$

where the principal uncertainty arises from the result of Gardner and Purcell. Since g_L equals 1, we can write

$$g_S = 2\mu_s = 2\,(1.001146 \pm 0.000012)$$

where μ_s is the spin magnetic moment of the electron in terms of μ_0.

The same result has subsequently been obtained by Beringer and Heald[19] who used a different experimental method involving, for atomic hydrogen, a microwave absorption technique and for the observation of the proton resonance frequency, the usual nuclear resonance technique. The primary result obtained by them was

$$g_J/g_P = -658.2181 \pm 0.0003$$

In view of the stated uncertainties and the possibility of differences in the internal diamagnetic shielding in different samples of mineral oil, the agreement is good. Because of the limited accuracy for the result $2g_L/g_P$, the value

of g_S/g_L is not affected, within its uncertainty, by the discrepancy, in the two results.

It is interesting to examine the ratio of g_S/g_L obtained by the sequence of experiments just described in light of the theoretical calculations of the electron moment. The result gives unambiguous evidence that the electron moment is anomalous and that the deviation of the moment from its nominal value is about $\alpha\mu_0/2\pi$. Karplus and Kroll[20] have calculated to a higher order the radiative correction to the spin moment of the electron and have found for the spin g value

$$g_S = 2\,(1 + \alpha/2\pi - 2.973\,\alpha^2/\pi^2) = 2\,(1.0011454)$$

The result of the experiment is in remarkable agreement with the calculation, especially since the uncertainty in the experiment is much greater than the discrepancy between the experimental and calculated values. The agreement offers conclusive evidence of the validity of the calculation to the order α and very strong support to the validity of the calculations to the order α^2. Thus the new procedures of quantum electrodynamics which have, perhaps, a questionable *a priori* validity are demonstrated to be, in practice, valid for the interpretation of certain observed phenomena and, therefore, useful in the exploration of other aspects of the behavior of matter.

1. I. I. Rabi, S. Millman, P. Kusch, and J. R. Zacharias, *Phys. Rev.*, 55 (1939) 526.
2. G. E. Uhlenbeck and S. Goudsmit, *Naturwiss.*, 13 (1925) 953; *Nature*, 117 (1926) 264.
3. H. A. Thomas, L. Driscoll, and J. A. Hipple, *Phys. Rev.*, 75 (1949) 902, 992; 78 (1950) 787.
4. P. Kusch, S. Millman, and I. I. Rabi, *Phys. Rev.*, 57 (1940) 765.
5. S. Millman and P. Kusch, *Phys. Rev.*, 60 (1941) 91.
6. J. E. Nafe and E. B. Nelson, *Phys. Rev.*, 73 (1948) 718.
7. G. Breit, *Phys. Rev.*, 72 (1947) 984.
8. P. Kusch and H. M. Foley, *Phys. Rev.*, 72 (1947) 1256; 73 (1948) 412; 74 (1948) 250.
9. T. C. Hardy and S. Millman, *Phys. Rev.*, 61 (1942) 459.
10. G. E. Becker and P. Kusch, *Phys. Rev.*, 73 (1948) 584.
11. A. K. Mann and P. Kusch, *Phys. Rev.*, 77 (1950) 435.
12. H. Taub and P. Kusch, *Phys. Rev.*, 75 (1949) 1481.
13. J. Schwinger, *Phys. Rev.*, 73 (1948) 416.

14. S. H. Koenig, A. G. Prodell, and P. Kusch, *Phys. Rev.*, 88 (1952) 191.
15. J. H. Gardner and E. M. Purcell, *Phys. Rev.*, 76 (1949) 1262.
16. A. G. Prodell and P. Kusch, *Phys. Rev.*, 88 (1952) 184.
17. J. B. Wittke and R. H. Dicke, *Phys. Rev.*, 96 (1954) 530.
18. P. Kusch, *Phys. Rev.*, 100 (1955) 1188.
19. R. Beringer and M. A. Heald, *Phys. Rev.*, 95 (1954) 1474.
20. R. Karplus and N. M. Kroll, *Phys. Rev.*, 77 (1950) 536.

Biography

Polykarp Kusch was born in Blankenburg, Germany, on the 26th January, 1911, the son of a clergyman. He has lived in the United States since 1912 and is a citizen of that country. He received his early education in the midwest of the United States. His original professional goal was in the field of chemistry, but soon after beginning his course of studies at the Case Institute of Technology, Cleveland, Ohio, his interest rapidly shifted to physics. In 1931 he received the B.S. degree in physics. He carried on his graduate study at the University of Illinois which awarded him the M.S. degree in 1933 and the Ph.D. degree in 1936. At Illinois he worked on problems in the field of optical molecular spectroscopy under the guidance of Professor F. Wheeler Loomis. He worked with Professor John T. Tate at the University of Minnesota in the field of mass spectroscopy during 1936–1937.

Since 1937 Kusch has been associated with the Department of Physics of Columbia University, New York City, except for interruptions engendered by World War II. These years were spent in research and development on microwave generators at the Westinghouse Electric Corporation, the Bell Telephone Laboratories and Columbia University. The experience was important not only in that it gave him knowledge of microwave methods, but also in that it suggested application of the special techniques of vacuum tube technology to a large range of problems in experimental physics.

Kusch has been a Professor of Physics at Columbia University since 1949. From his first days at Columbia, he has been intimately associated with Professor I. I. Rabi in his programme of research on atomic, molecular and nuclear properties and phenomena by the method of molecular beams. The direction in which his own research has been directed has been greatly influenced by this long association. His research has dealt principally with the small details of the interactions of the constituent particles of atoms and of molecules with each other and with externally applied fields. The establishment of the reality of the anomalous magnetic moment of the electron and the precision determination of its magnitude was part of an intensive programme of postwar research with atomic and molecular beams. Later,

he has also become interested in problems in chemical physics to whose experimental study he has applied the molecular beams technique.

Professor Kusch has been awarded honorary Sc.D. degrees of the Case Institute of Technology, the Ohio State University, the University of Illinois and Colby College. He was elected to the membership in the National Academy of Sciences (USA) in 1956.

In recent years he is increasingly concerned with problems of education, especially that of educating the young to understand a civilization strongly affected by the knowledge of science and by the techniques that result from this knowledge.

Kusch married Edith Starr McRoberts; they had three daughters. His wife died in 1959 and he was married to Betty Pezzoni in 1960.

Physics 1956

WILLIAM SHOCKLEY

JOHN BARDEEN

WALTER HOUSER BRATTAIN

«for their researches on semiconductors and their discovery of the transistor effect»

Physics 1956

Presentation Speech by Professor E. G. Rudberg, member of the Nobel Committee for Physics

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

In these days 250 years have elapsed since Benjamin Franklin was born: the printer and educator, the statesman, the pioneer in the field of electricity. It was Franklin who strung a high-tension line from a thundercloud to a green pasture in idyllically rural Philadelphia. He showed that the cloud held electric energy. A kite drew energy out of the cloud. The kite string was drenched by the rain and therefore conducted the charge down to a key, which gave off sparks when approached too closely. Franklin had tied one end of a silk ribbon to the key; he clutched the other end of the ribbon as he stooped under a cowshed to keep his silk insulator dry.

A conductor and an insulator was what Franklin needed for his power line. Electrical engineering would have been unthinkable today if Nature had not presented the material in these two extreme classes, metallic conductors and insulators. Mobile carriers of charge are almost entirely lacking in an insulator, but a good conductor has plenty of them, about one for each atom. As early as 100 years ago they carried the current in the first Atlantic cable from the Old World to the New – in a fraction of a second. A group of charged carriers enters at the European end, and immediately afterwards carriers emerge from the American end – but not the same group. Over the entire length of the cable, carriers are standing tightly packed. The emigrants must push to make room for themselves at the very entrance. This push darts as a shock with the speed of light down the long line of carriers finally ejecting those that are standing next to the exit in America. Charge is therefore transported with lightning speed, although each carrier only moves a short distance. In the old days, carriers were thought to be of two kinds, positive and negative, moving in opposite directions. Franklin held that only one kind was needed. Franklin's contention was supported by the great discoveries around the year 1900. The carriers in metallic conductors are electrons, and they all carry the same negative charge.

If the Easter pilgrims in Piazza San Pietro were to represent the carriers in a metal, then an insulator would resemble the Antarctic with one solitary

traveller. In the abundance of carriers there is an enormous gap between conductors and insulators. In this gap it is now possible to place the semiconductors, with carriers about as numerous as the longshoremen in a harbour when a loaded freighter has just arrived. The semiconductors now in use are artificial products made from elements such as germanium or silicon. The pure element has very few carriers. Through small additions of certain contaminants, however, it is possible to alter the supply of carriers. Every atom of phosphorus, forced as a lodger on silicon, donates one carrier to the house, a negative electron. A few parts in 100,000 make a good semiconductor. Still more remarkable is that a guest atom of boron provides a carrier of the opposite kind – positive. This the guest manages to accomplish by stealing an electron which his host, silicon, had kept locked up. Where that electron was, a hole is now left. This hole can migrate in the semiconductor, and it then acts as a carrier of positive charge.

It is possible to have both holes and electrons as carriers in a semiconductor at the same time. Donors and thieves are lodged in such proportions that one kind of carrier, or the opposite kind, will prevail. Much of the technical importance of semiconductors stems from the interplay of holes and electrons. The idea of two kinds of carriers is contrary to Franklin's views. This idea was put forward in the 1930's, at a time when rectifiers based on semiconductors began to find important uses. Attempts were made to control these rectifiers by means of an extra electrode, just as a radio valve is controlled by the grid – without success. Finally, in 1948, the discovery of transistor action gave Shockley, Bardeen, and Brattain the key to the control mechanism and, in addition, a new weapon for tackling the semiconductor problems.

The description must now borrow a picture from the classical books of adventure. To place a negative electrode against a semiconductor with negative carriers – this is like bringing a ship up to a quay in the Orient, with the yellow flag of the plague hoisted. The place becomes deserted by its carriers. Unloading – current – is blocked. But exchange that negative flag of pestilence for a positive sign and the carriers will return, the contact becoming conducting. Electrically this is called rectification. In those seafaring tales it was perhaps possible to induce the carriers to return, without striking the flag, merely by throwing some gold coins on the quay, thus positively destroying the insulation. It is possible to destroy the blockade in the semiconductor in a similar fashion by throwing in some positive holes around which the negative carriers will gather. This is transistor action. It is a fine

thing that the carriers' strike can be broken up by rather few holes, which do not cost much energy. Thus the current in the rectifier is controlled through the injection of holes. A transistor functions much like a radio valve. But it is smaller, and it does not require current to heat a filament. Hearing aids, computing machines, telephone stations and many others are in need of just such a device.

The physicists at Murray Hill decided to map out that region, poor in carriers, near a negative electrode, using a movable probe at the surface of the semiconductor. This is done in the same fashion as electric prospecting for ore, but the scale is a different one. Bardeen and Brattain moved their tiny probe under the microscope, using a micrometer screw. When the probe was made positive quite close to the electrode they found that the blockade was lifted. The probe acted as an injector of holes. Shockley and his collaborators hastened to utilize this injector in a series of ingeniously conceived experiments, which then disclosed many properties of holes: how fast they travel, how long they live and other characteristics. With new tools such as these, semiconductor physics is today a seething field of research.

From Philadelphia's old pasture to today's Murray Hill is not many miles – but 200 years. Evidently there is more than the geographical proximity that connects Franklin's work with the discoveries of his latter-day countrymen.

Doctor Shockley, Doctor Bardeen, Doctor Brattain. The summit of Everest was reached by a small party of ardent climbers. Working from an advance base, they succeeded. More than a generation of mountaineers had toiled to establish that base. Your assault on the semiconductor problem was likewise launched from a high-altitude camp, contributed by many scientists. Yours, too, was a supreme effort – of foresight, ingenuity and perseverence, excercised individually and as a team. Surely, supreme joy befalls the man to whom those breathtaking vistas from the summit unfold. You must have felt it, overwhelmingly. This joy is now shared by those who laboured at the base. Shared, too, is the challenge of untrodden territory, now seen for the first time, calling for a new scientific attack.

Thus salutes you, Nobel Laureates, the Royal Academy of Sciences.

And now, my solemn duty, nay, my treasured privilege: to invite you to receive your award from the hands of His Majesty the King.

JOHN BARDEEN

Semiconductor research leading to the point contact transistor

Nobel Lecture, December 11, 1956

Introduction

In this lecture we shall attempt to describe the ideas and experiments which led to the discovery of the transistor effect as embodied in the point-contact transistor. Some of the important research done subsequent to the discovery will be described in the following lectures by Shockley and Brattain. As we shall see, the discovery was but a step along the road of semiconductor research to which a great many people in different countries have contributed. It was dependent both on the sound theoretical foundation largely built up during the thirties and on improvement and purification of materials, particularly of germanium and silicon, in the forties. About half of the lecture will be devoted to an outline of concepts concerning electrical conduction in semiconductors and rectification at metal-semiconductor contacts as they were known at the start of our research program.

The discovery of the transistor effect occurred in the course of a fundamental research program on semiconductors initiated at the Bell Telephone Laboratories in early 1946. Semiconductors was one of several areas selected under a broad program of solid-state research, of which S. O. Morgan and W. Shockley were co-heads. In the initial semiconductor group, under the general direction of Shockley, were W. H. Brattain, concerned mainly with surface properties and rectification, G. L. Pearson, concerned with bulk properties, and the writer, interested in theoretical aspects of both. Later a physical chemist, R. B. Gibney, and a circuit expert, H. R. Moore, joined the group and made important contributions, particularly to chemical and instrumentation problems, respectively.

It is interesting to note that although Brattain and Pearson had had considerable experience in the field prior to the war, none of us had worked on semiconductors during the war years. We were able to take advantage of the important advances made in that period in connection with the development of silicon and germanium detectors and at the same time have a

fresh look at the problems. Considerable help was obtained from other groups in the Laboratories which were concerned more directly with wartime developments. Particular mention should be made of J. H. Scaff, H. C. Theuerer and R. S. Ohl.

The general aim of the program was to obtain as complete an understanding as possible of semiconductor phenomena, not in empirical terms, but on the basis of atomic theory. A sound theoretical foundation was available from work done during the thirties:

(1) Wilson's quantum mechanical theory[1], based on the energy band model, and describing conduction in terms of excess electrons and holes. It is fundamental to all subsequent developments. The theory shows how the concentration of carriers depends on the temperature and on impurities.

(2) Frenkel's theories of certain photoconductive phenomena[2] (change of contact potential with illumination and the photomagneto electric effect) in which general equations were introduced which describe current flow when non-equilibrium concentrations of both holes and conduction electrons are present. He recognized that flow may occur by diffusion in a concentration gradient as well as by an electric field.

(3) Independent and parallel developments of theories of contact rectification by Mott[3], Schottky[4] and Davydov[5]. The most complete mathematical theories were worked out by Schottky and his co-worker, Spenke.

Of great importance for our research program was the development during and since the war of methods of purification and control of the electrical properties of germanium and silicon. These materials were chosen for most of our work because they are well-suited to fundamental investigations with the desired close coordination of theory and experiment. Depending on the nature of the chemical impurities present, they can be made to conduct by either excess electrons or holes.

Largely because of commercial importance in rectifiers, most experimental work in the thirties was done on copper oxide (Cu_2O) and selenium. Both have complex structures and conductivities which are difficult to control. While the theories provided a good qualitative understanding of many semiconductor phenomena, they had not been subjected to really convincing quantitative checks. In some cases, particularly in rectification, discrepancies between experiment and theory were quite large. It was not certain whether the difficulties were caused by something missing in the theories or by the fact that the materials used to check the theories were far from ideal.

In the U.S.A., research on germanium and silicon was carried out during

the war by a number of university, government and industrial laboratories in connection with the development of point-contact or « cat's whisker » detectors for radar. Particular mention should be made of the study of germanium by a group at Purdue University working under the direction of K. Lark-Horovitz and of silicon by a group at the Bell Telephone Laboratories. The latter study was initiated by R. S. Ohl before the war and carried out subsequently by him and by a group under J. H. Scaff. By 1946 it was possible to produce relatively pure polycrystalline materials and to control the electrical properties by introducing appropriate amounts of donor and acceptor impurities. Some of the earliest work (1915) on the electrical properties of germanium and silicon was done in Sweden by Prof. C. Benedicks.

Aside from intrinsic scientific interest, an important reason for choosing semiconductors as a promising field in which to work, was the many and increasing applications in electronic devices, which, in 1945, included diodes, varistors and thermistors. There had long been the hope of making a triode, or an amplifying device with a semiconductor. Two possibilities had been suggested. One followed from the analogy between a metal semiconductor rectifying contact and a vacuum-tube diode. If one could somehow insert a grid in the space-charge layer at the contact, one should be able to control the flow of electrons across the contact. A major practical difficulty is that the width of the space-charge layer is typically only about 10^{-4} cm. That the principle is a sound one was demonstrated by Hilsch and Pohl[6], who built a triode in an alkali-halide crystal in which the width of the space-charge layer was of the order of one centimeter. Because amplification was limited to frequencies of less than one cycle per second, it was not practical for electronic applications.

The second suggestion was to control the conductance of a thin film or slab of semiconductor by application of a transverse electric field (called the *field effect*). In a simple form, the slab forms one plate of a parallel plate condenser, the control electrode being the other plate. When a voltage is applied across the condenser, charges are induced in the slab. If the induced charges are mobile carriers, the conductance should change with changes of voltage on the control electrode. This form was suggested by Shockley; his calculations indicated that, with suitable geometry and materials, the effect should be large enough to produce amplification of an a.c. signal[7].

Point-contact and junction transistors operate on a different principle than either of these two suggestions, one not anticipated at the start of the program. The transistor principle, in which both electrons and holes play a role,

was discovered in the course of a basic research program on surface properties.

Shockley's field-effect proposal, although initially unsuccessful, had an important bearing on directing the research program toward a study of surface phenomena and surface states. Several tests which Shockley carried out at various times with J. R. Haynes, H. J. McSkimin, W. A. Yager and R. S. Ohl, using evaporated films of germanium and silicon, all gave negative results. In analyzing the reasons for this failure, it was suggested[8] that there were states for electrons localized at the surface, and that a large fraction of the induced charge was immobilized in these states. Surface states also accounted for a number of hitherto puzzling features of germanium and silicon point-contact diodes.

In addition to the possibility of practical applications, research on surface properties appeared quite promising from the viewpoint of fundamental science. Although surface states had been predicted as a theoretical possibility, little was known about them from experiment. The decision was made, therefore, to stress research in this area. The study of surfaces initiated at that time (1946) has been continued at the Bell Laboratories and is now being carried out by many other groups as well[9].

It is interesting to note that the field effect, originally suggested for possible value for a device, has been an extremely fruitful tool for the fundamental investigation of surface states. Further, with improvements in semiconductor technology, it is now possible to make electronic amplifiers with high gain which operate on the field-effect principle.

Before discussing the research program, we shall give first some general background material on conduction in semiconductors and metal-semiconductor rectifying contacts.

Nature of conduction in semiconductors

An electronic semiconductor is typically a valence crystal whose conductivity depends markedly on temperature and on the presence of minute amounts of foreign impurities. The ideal crystal at the absolute zero is an insulator. When the valence bonds are completely occupied and there are no extra electrons in the crystal, there is no possibility for current to flow. Charges can be transferred only when imperfections are present in the electronic structure, and these can be of two types: *excess electrons* which do not

fit into the valence bonds and can move through the crystal, and *holes*, places from which electrons are missing in the bonds, which also behave as mobile carriers. While the excess electrons have the normal negative electronic charge, $-e$, holes have a positive charge, $+e$. It is a case of two negatives making a positive; a missing negative charge is a positive defect in the electron structure.

The bulk of a semiconductor is electrically neutral; there are as many positive charges as negative. In an *intrinsic* semiconductor, in which current carriers are created by thermal excitation, there are approximately equal numbers of excess electrons and holes. Conductivity in an *extrinsic* semiconductor results from impurity ions in the lattice. In n-type material, the negative charge of the excess electrons is balanced by a net positive space charge of impurity ions. In p-type, the *positive* charge of the holes is balanced by negatively charged impurities. Foreign atoms which can become positively charged on introduction to the lattice are called *donors*; atoms which become negatively ionized are called *acceptors*. Thus donors make a semiconductor n-type, acceptors p-type. When both donors and acceptors are present, the conductivity type depends on which is in excess. Mobile carriers then balance the *net* space charge of the impurity ions. Terminology used is listed in the table below:

Table 1.

Designation of conductivity type		*Majority carrier*	*Dominant impurity ion*
n-type	excess	electron (n/cm^3)	donor
p-type	defect	hole (p/cm^3)	acceptor

These ideas can be illustrated quite simply for silicon and germanium, which, like carbon, have a valence of four and lie below carbon in the Periodic Table. Both crystallize in the diamond structure in which each atom is surrounded tetrahedrally by four others with which it forms bonds. Carbon in the form of a diamond is normally an insulator; the bond structure is complete and there are no excess electrons. If ultraviolet light falls on diamond, electrons can be ejected from the bond positions by the photoelectric effect. Excess electrons and holes so formed can conduct electricity; the crystal becomes photoconductive.

The energy required to free an electron from a bond position so that it and the hole left behind can move the crystal, is much less in silicon and germanium than for diamond. Appreciable numbers are released by thermal excitations at high temperatures; this gives intrinsic conductivity.

Impurity atoms in germanium and silicon with more than four valence electrons are usually donors, those with less than four acceptors. For example, Group V elements are donors, Group III elements acceptors. When an arsenic atom, a Group V element, substitutes for germanium in the crystal, only four of its valence electrons are required to form the bonds. The fifth is only weakly held by the forces of Coulomb attraction, greatly reduced by the high dielectric constant of the crystal. The energy required to free the extra electron is so small that the arsenic atoms are completely ionized at room temperature. Gallium, a typical Group III acceptor, has only three valence electrons. In order to fill the four bonds, Ga picks up another electron and enters the crystal in the form of a negative ion, Ga^-. The charge is balanced by a free hole.

While some of the general notions of excess and defect conductivity, donors and acceptors, go back earlier, Wilson[1] was the first to formalize an adequate mathematical theory in terms of the band picture of solids. The band picture itself, first applied to metals, is a consequence of an application of quantum mechanics to the motion of electrons in the periodic potential field of a crystal lattice. Energy levels of electrons in atoms are discrete. When the atoms are combined to form a crystal, the allowed levels form continuous bands. When a band is completely occupied, the net current of all of the electrons in the band is zero. Metals have incompletely filled bands. In insulators and semiconductors, there is an energy gap between the highest filled band and the next higher allowed band of levels, normally unoccupied.

The relations are most simply illustrated in terms of an energy-level diagram of the crystal. In Fig. 1 is shown a schematic energy-level diagram of an intrinsic semiconductor. Electrons taking part in the chemical bonds form a continuous band of levels called the valence band. Above these is an energy gap in which there are no allowed levels in the ideal crystal, and then another continuous band of levels called the conduction band. The energy gap, E_G, is the energy required to free an electron from the valence bonds. Excess, or conduction, electrons have energies in the lower part of the conduction band. The very lowest state in this band, E_C, corresponds to an electron at rest, the higher states to electrons moving through the crystal with

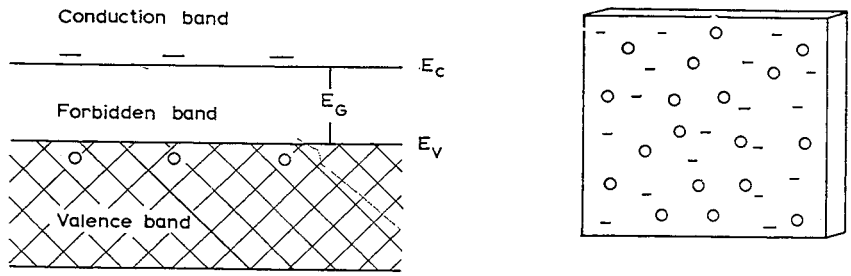


Fig. 1. Energy-level diagram of an intrinsic semiconductor. There is a random distribution of electrons and holes in equal numbers.

additional energy of motion. Holes correspond to states near the top of the valence band, E_V, from which electrons are missing. In an intrinsic semiconductor, electrons and holes are created in equal numbers by thermal excitation of electrons from the valence to the conduction band, and they are distributed at random through the crystal.

In an n-type semiconductor, as illustrated in Fig. 2a, there is a large number of electrons in the conduction band and very few holes in the valence band. Energy levels corresponding to electrons localized around Group V donor impurity atoms are typically in the forbidden gap and a little below the conduction band. This means that only a small energy is required to ionize the donor and place the electron removed in the conduction band. The charge of the electrons in the conduction band is compensated by the positive space charge of the donor ions. Levels of Group III acceptors (Fig. 2b) are a little above the valence band. When occupied by thermal excitation of electrons from the valence band, they become negatively charged. The space charge of the holes so created is compensated by that of the negative acceptor ions.

Occupancy of the levels is given by the position of the Fermi level, E_F. The probability, f, that a level of energy E is occupied by an electron is given by the Fermi-Dirac function:

$$f = \frac{1}{1 + \exp\,(E - E_F)/kT}$$

The energy gap in a semiconductor is usually large compared with thermal energy, kT ($\sim$ 0.025 eV at room temperature), so that for levels well above E_F one can use the approximation

$$f \simeq \exp\left[-(E - E_F)/kT\right]$$

For levels below E_F, it is often more convenient to give the probability

$$f_p = 1 - f = \frac{1}{1 + \exp\left[(E_F - E)/kT\right]}$$

that the level is unoccupied, or « occupied by a hole ». Again, for levels well below E_F,

$$f_p \simeq \exp\left[-(E_F - E)/kT\right]$$

The expressions for the total electron and hole concentrations (number per unit volume), designated by the symbols n and p respectively, are of the form

$$n = N_C \exp\left[-(E_C - E_F)/kT\right]$$
$$p = N_V \exp\left[-(E_F - E_V)/kT\right]$$

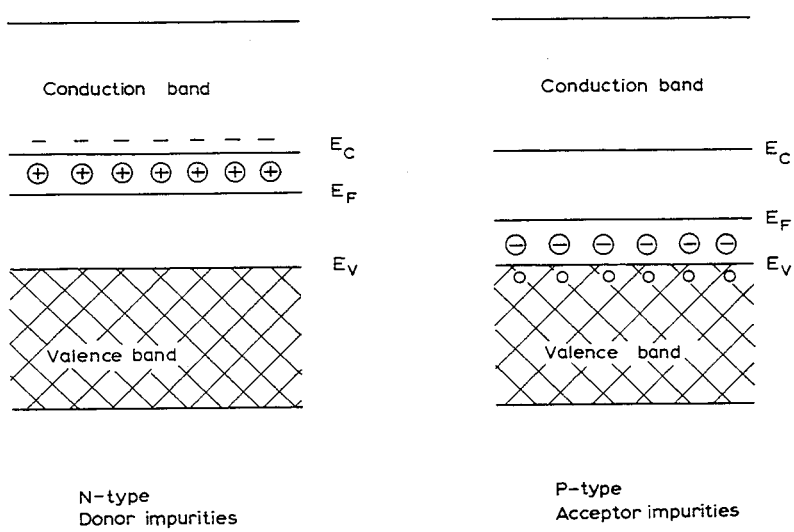


Fig. 2. Energy-level diagrams for n- and p-type semiconductors.

where N_C and N_V vary slowly with temperature compared with the exponential factors. Note that the product np is independent of the position of the Fermi level and depends only on the temperature:

$$np = n_i^2 = N_C N_V \exp\left[-(E_C - E_V)/kT\right] = N_C N_V \exp\left[-E_G/kT\right]$$

Here n is the concentration in an intrinsic semiconductor for which $n = p$.

In an n-type semiconductor, the Fermi level is above the middle of the gap, so that $n \gg p$. The value of n is fixed by the concentration of donor ions, N_d^+, so that there is electrical neutrality:

$$n - p = N_d^+$$

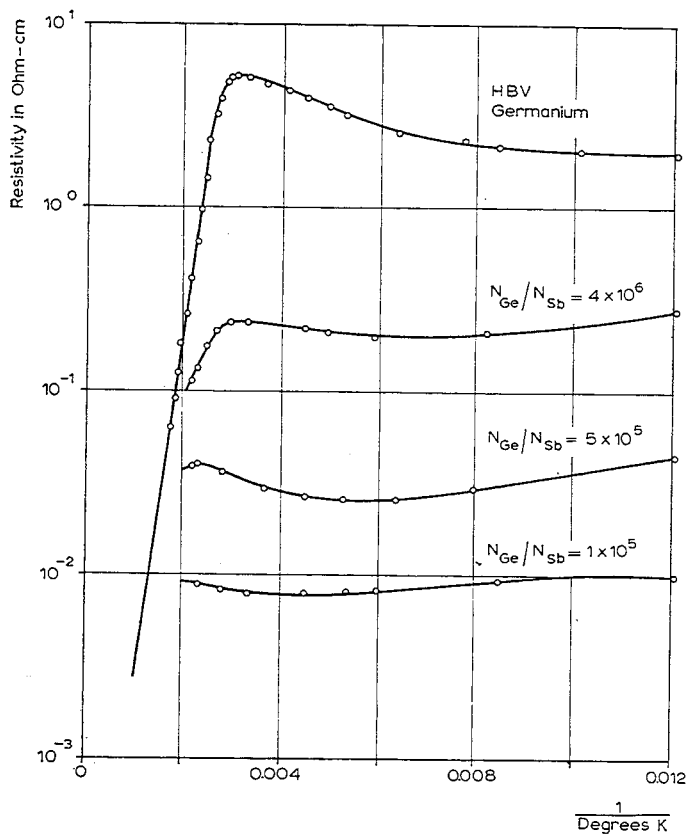


Fig. 3. Conductivity *vs.* $1/T$ for germanium with antimony added as a donor impurity.

The minority carrier concentration, p, increases rapidly with temperature and eventually a temperature will be reached above which n and p are both large compared with N_d^+ and the conduction is essentially intrinsic. Correspondingly in a p-type semiconductor, in which there are acceptor ions, $p \gg n$, and the Fermi level is below the center of the gap.

The Fermi level is equivalent to the chemical potential of the electrons. If two conductors are electrically connected together so that electrons can be transferred, the relative electrostatic potentials will be adjusted so that the Fermi levels of the two are the same. If the n- and p-type materials of Fig. 2 are connected, a small number of electrons will be transferred from the n-type to the p-type. This will charge the p-type negatively with respect to the n-type and raise the electrostatic potential energy of the electrons accordingly. Electron transfer will take place until the energy levels of the p-type material are raised relative to those of the n-type by the amount required to make the Fermi levels coincide.

The amount of impurity required to make significant changes in the conductivity of germanium or silicon is very small. There is given in Fig. 3 a plot, on a log scale, of the resistivity *vs.* $1/T$ for specimens of germanium with varying amounts of antimony, a donor impurity. This plot is based on some measurements made by Pearson several years ago[11]. The purest specimens available at that time had a room temperature resistivity of about 10–20 ohm cm, corresponding to about one donor atom in 10^8 germanium atoms. This material (H.B.V.) is of the sort which was used to make germanium diodes which withstand a high voltage in the reverse direction (*H*igh *B*ack *V*oltage) and also used in the first transistors. The purest material available now corresponds to about one donor or acceptor in 10^{10}. The resistivity drops, as illustrated, with increasing antimony concentration; as little as one part in 10^7 makes a big difference. All specimens approach the intrinsic line corresponding to pure germanium at high temperatures.

Conduction electrons and holes are highly mobile, and may move through the crystal for distances of hundreds or thousands of the interatomic distance, before being scattered by thermal motion or by impurities or other imperfections. This is to be understood in terms of the wave property of the electron; a wave can travel through a perfect periodic structure without attenuation. In treating acceleration in electric or magnetic fields, the wave aspect can often be disregarded, and electrons and holes thought of as classical particles with an effective mass of the same order, but differing from the ordinary electron mass. The effective mass is often anisotropic, and dif-

ferent for different directions of motion in the crystal. This same effective mass picture can be used to estimate the thermal motion of the gas of electrons and holes. Average thermal velocities at room temperature are of the order of 10^7 cm/sec.

Scattering can be described in terms of a mean free path for the electrons and holes. In relatively pure crystals at ordinary temperatures, scattering occurs mainly by interaction with the thermal vibrations of the atoms of the crystal. In less pure crystals, or in relatively pure crystals at low temperatures, the mean free path may be determined by scattering by impurity atoms. Because of the scattering, the carriers are not uniformly accelerated by an electric field, but attain an average drift velocity proportional to the field. Ordinarily the drift velocity is much smaller than the average thermal velocity. Drift velocities may be expressed in terms of the mobilities, μ_n and μ_p of the electrons and holes respectively*.

In an electric field E,

$$(V_d)_n = -\mu_n E$$
$$(V_d)_p = \mu_p E$$

Because of their negative charge, conduction electrons drift oppositely to the field. Values for pure germanium at room temperature are $\mu_n = 3{,}800$ cm^2/volt sec; $\mu_p = 1{,}800$ cm^2/volt sec. This means that holes attain a drift velocity of 1,800 cm/sec in a field of one volt/cm.

Expressions for the conductivity are:

$$\text{n-type:} \quad \sigma_n = ne\mu_n$$
$$\text{p-type:} \quad \sigma_p = pe\mu_p$$
$$\text{intrinsic:} \quad \sigma = ne\mu_n + pe\mu_p$$

It is not possible to determine n and μ_n separately from measurements of the conductivity alone. There are several methods to determine the mobility; one which has been widely used is to measure the Hall coefficient in addition to the conductivity. As part of the research program at the Bell Laboratories, Pearson and Hall made resistivity measurements over a wide range of temperatures of silicon containing varying amounts of boron (a Group III ac-

* A subscript n (referring to *n*egative charge) is used for conduction electrons, p (*p*ositive) for holes.

ceptor) and of phosphorus (a Group V donor). Analysis of the data[10] gave additional confirmation of the theory we have outlined. Similar measurements on germanium were made about the same time by Lark-Horovitz and co-workers, and more recently more complete measurements on both materials have been made by other groups. The result of a large amount of experimental and theoretic work has been to confirm the Wilson model in quantitative detail.

Carriers move not only under the influence of an electric field, but also by diffusion; the diffusion current is proportional to the concentration gradient. Expressions for the particle current densities of holes and electrons, respectively, are

$$j_p = p\mu_p E - D_p \operatorname{grad} p$$
$$j_n = n\mu_n E - D_n \operatorname{grad} n$$

Einstein has shown that mobilities and diffusion coefficients are related:

$$\mu = \frac{e}{kT} D$$

where k is Boltzmann's constant. Diffusion and conduction currents both play an important role in the transistor.

The diffusion term was first considered by Wagner in his theory of oxidation of metals. The equations were worked out more completely by Frenkel[2] in an analysis of the diffusive flow which occurs when light is absorbed near one face of a slab, as shown schematically in Fig. 4. The light quanta raise

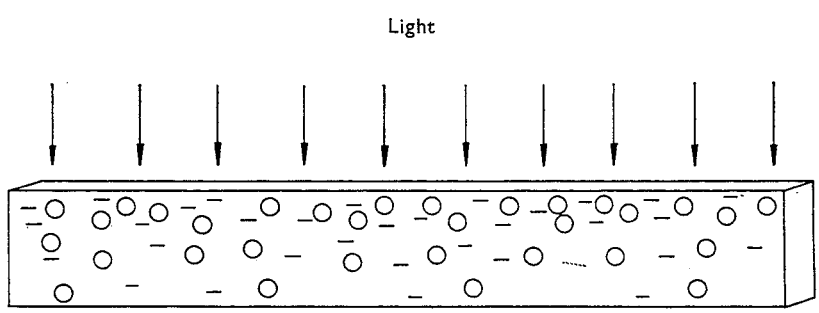


Fig. 4. Schematic diagram of diffusive flow of electrons and holes created near the surface by absorption of light.

electrons from the valence to the conduction bands, creating conduction electrons and holes in equal numbers. These diffuse toward the interior of the slab. Because of recombination of conduction electron and hole pairs, the concentration drops as the diffusion occurs. Frenkel gave the general equations of flow when electrons and holes are present in attempting to account for the Dember effect (change in contact potential with light) and the photomagnetoelectric (PME) effect. The latter is a voltage analogous to a Hall voltage observed between the ends of a slab in a transverse magnetic field (perpendicular to the paper in the diagram). The Dember voltage was presumed to result from a difference of mobility, and thus of diffusion coefficient, between electrons and holes. Electrical neutrality requires that the concentrations and thus the concentration gradients be the same. Further, under steady-state conditions the flow of electrons to the interior must equal the flow of holes, so that there is no net electrical current. However, if D_n is greater than D_p, the diffusive flow of electrons would be greater than that of holes. What happens is that an electric field, E, is introduced which aids holes and retards the electrons so as to equalize the flows. The integral of E gives a voltage difference between the surface and the interior, and thus a change in contact potential. As we will mention later, much larger changes in contact potential with light may come from surface barrier effects.

Contact rectifiers

In order to understand how a point-contact transistor operates, it is necessary to know some of the features of a rectifying contact between a metal and semiconductor. Common examples are copper-oxide and selenium rectifiers and germanium and silicon point-contact diodes which pass current much more readily for one direction of applied voltage than the opposite. We shall follow Schottky's picture[4], and use as an illustration a contact to an n-type semiconductor. Similar arguments apply to p-type rectifiers with appropriate changes of sign of the potentials and charges. It is most convenient to make use of an energy-level diagram in which the changes in energy bands resulting from changes in electrostatic potential are plotted along a line perpendicular to the contact, as in Fig. 5. Rectification results from the potential energy barrier at the interface which impedes the flow of electrons across the contact.

The Fermi level of the metal is close to the highest of the normally oc-

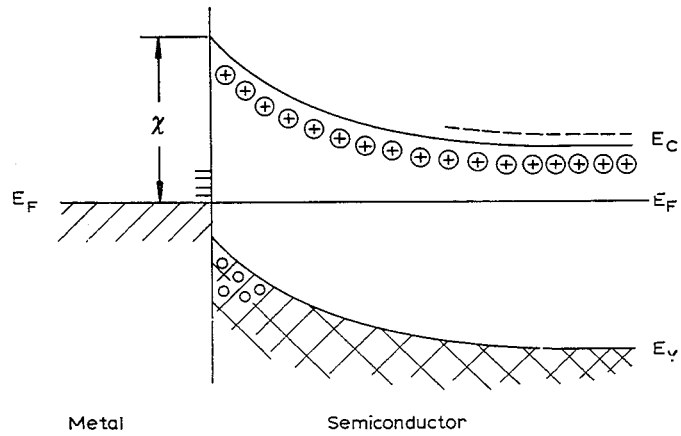


Fig. 5. Equilibrium energy-level diagram for a metal-semiconductor rectifying contact along a line perpendicular to the interface. Variations in the energy bands of the semiconductor result from changes in electrostatic potential due to the layer of uncompensated space-charge. The overall change in potential from the surface to the interior is such as to bring the Fermi level in the interior of the semiconductor into coincidence with that of the metal. In this example, there is an inversion from n-type conductance in the bulk to p-type at the surface.

cupied levels of the conduction band. Because of the nature of the metal-semiconductor interface layers, a relatively large energy, χ, perhaps of the order of 0.5 eV, is required to take an electron from the Fermi level of the metal and place it in the conduction band in the semiconductor. In the interior of the semiconductor, which is electrically neutral, the position of the Fermi level relative to the energy bands is determined by the concentration of conduction electrons, and thus of donors. In equilibrium, with no voltage applied, the Fermi levels of the metal and semiconductor must be the same. This is accomplished by a region of space charge adjacent to the metal in which there is a variation of electrostatic potential, and thus of potential energy of the electron, as illustrated.

In the bulk of the semiconductor there is a balance between conduction electrons and positive donors. In the barrier region which is one of high potential energy for electrons, there are few electrons in the conduction band. The uncompensated space charge of the donors is balanced by a negative charge at the immediate interface. It is these charges, in turn, which produce the potential barrier. The width of the space-charge region is typically of the order of 10^{-5} to 10^{-4} cm.

When a voltage is applied, most of the drop occurs across the barrier layer. The direction of easy flow is that in which the semiconductor is negative relative to the metal. The bands are raised, the barrier becomes narrower, and electrons can flow more easily from the semiconductor to the metal. In the high resistance direction, the semiconductor is positive, the bands are lowered relative to the metal, and the barrier is broadened. The current of electrons flowing from the metal is limited by the energy barrier, χ, which must be surmounted by thermal excitation.

If χ is sufficiently large, the Fermi level at the interface may be close to the valence band, implying an inversion from n-type conductivity in the bulk to p-type near the contact. The region of hole conduction is called, following Schottky, an inversion layer. An appreciable part of the current flow to the contact may then consist of minority carriers, in this case holes. An important result of the research program at the Bell Laboratories after the war was to point out the significance of minority carrier flow.

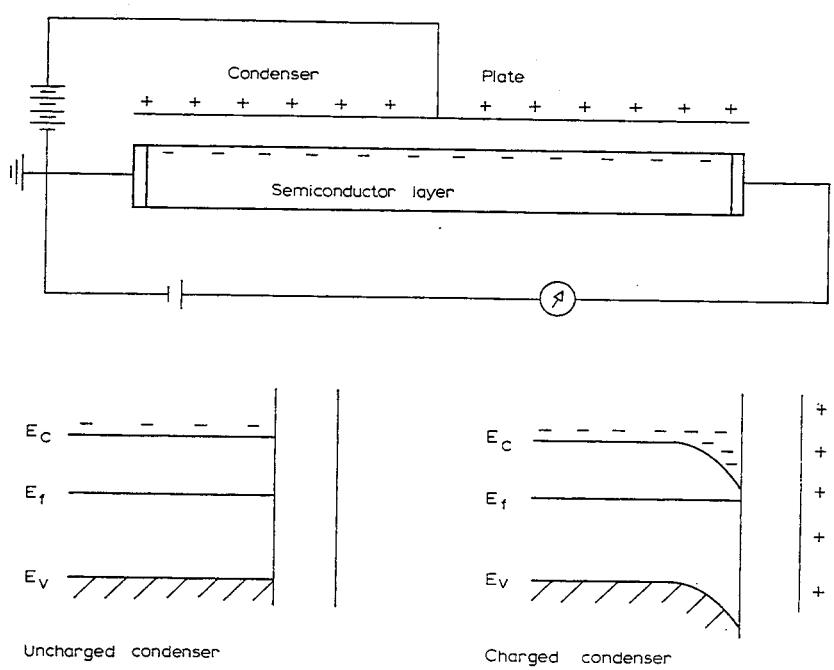


Fig. 6. Schematic diagram of a field-effect experiment for an n-type semiconductor with no surface states.

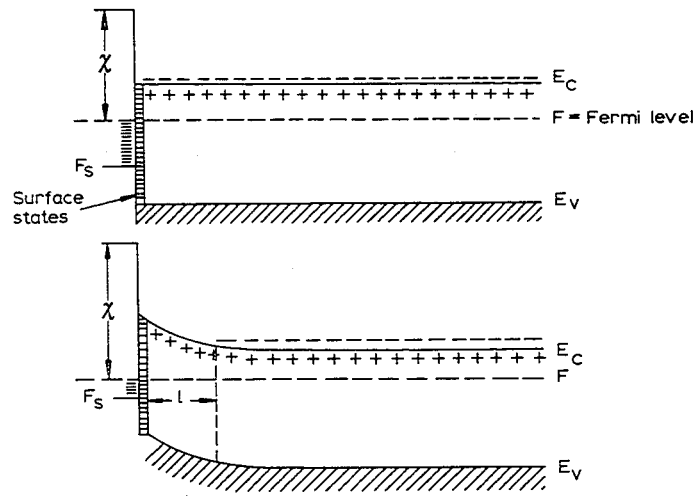


Fig. 7. Formation of a space-charge barrier layer at the free surface of a semiconductor

Experiments on surface states

We have mentioned in the introduction that the negative result of the field-effect experiment was an important factor in suggesting the existence of surface states on germanium and silicon, and directing the research program toward a study of surface properties. As is shown in Fig. 6, the experiment consists of making a thin film or slab one plate of a parallel plate condenser and then measuring the change in conductance of the slab with changes in voltage applied across the condenser. The hypothetical case illustrated is an n-type semiconductor with no surface states. When the field plate is positive, the negative charge induced on the semiconductor consists of added electrons in the conduction band. The amount of induced charge can be determined from the applied voltage and measured capacity of the system. If the mobility is known, the expected change in conductance can be calculated readily.

When experiments were performed on evaporated films of germanium and silicon, negative results were obtained; in some cases the predicted effect was more than one thousand times the experimental limit of detection. Analysis indicated that a large part of the discrepancy, perhaps a factor of 50 to 100, came from the very low mobility of electrons in the films as compared with bulk material. The remaining was attributed to shielding by surface states.

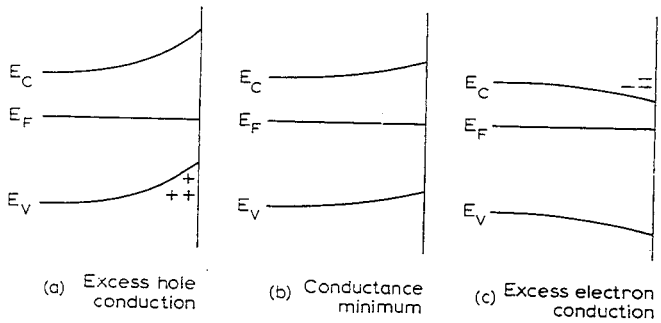


Fig. 8. Types of barrier layers which may exist at the free surface of an n-type semiconductor: (*a*) excess conductance from an inversion layer of p-type conductivity; (*b*) near the minimum surface conductance; (*c*) excess conductance from an accumulation layer of electrons.

It was predicted that if surface states exist, a barrier layer of the type found at a metal contact might be found at the free surface of a semiconductor. The formation of such a layer is illustrated schematically in Fig. 7. Occupancy of the surface levels is determined by the position of the Fermi level at the surface. In the illustration, it is presumed that the distribution of surface states is such that the states themselves would be electrically neutral if the Fermi level crossed at the position F_S relative to the bands. If there is no surface barrier, so that the Fermi level crosses the surface above F_S, there are excess electrons and a net negative charge in the surface states. When the surface as whole is neutral, a barrier layer is formed such that the positive charge in the layer is compensated by the negative surface states charge. If the density of surface states is reasonably high, sufficient negative charge is obtained with the Fermi level crossing only slightly above F_S.

Types of barriers which may exist at the surface of an n-type semiconductor are illustrated in Fig. 8. On the left (*a*) the energy bands are raised at the surface so as to bring the valence band close to the Fermi level. An inversion layer of opposite conductivity type is formed, and there is excess conductance from mobile holes in the layer. Negative charge on the surface proper is balanced by the charge of holes and of fixed donor ions in the barrier region. In (*b*) the bands are raised at the surface, but not enough to form a barrier layer. The surface resistance is near a maximum. In (*c*), the bands bend down so as to form an *accumulation* layer of excess electron conductance near the surface. The charge on the surface proper is now positive, and is balanced by the negative charge of the excess electrons in the layer.

The postulated existence of surface states and surface barrier layers on the free surface of germanium and silicon accounted for several properties of germanium and silicon which had hitherto been puzzling[8]. These included (1) lack of dependence of rectifier characteristics on the work function of the metal contact, (2) current voltage characteristics of a contact made with two pieces of germanium, and (3) the fact that there was found little or no contact potential difference between n- and p-type germanium and between n- and p-type silicon.

While available evidence for surface states was fairly convincing, it was all of an indirect nature. Further, none of the effects gave any evidence about the height of the surface barrier and of the distribution of surface states. A number of further experiments which might yield more concrete evidence about the surface barrier was suggested by Shockley, Brattain and myself. Shockley predicted that a difference in contact potential would be found between n- and p-type specimens with large impurity concentration. A systematic study of Brattain and Shockley[12] using silicon specimens with varying amounts of donor and acceptor impurities showed that this was true, and an estimate was obtained for the density of surface states. Another experiment which indicated the presence of a surface barrier was a measurement of the change in contact potential with illumination of the surface. This is just the Dember effect, which Frenkel had attempted to account for by the difference in mobilities of the electrons and holes generated by the light and

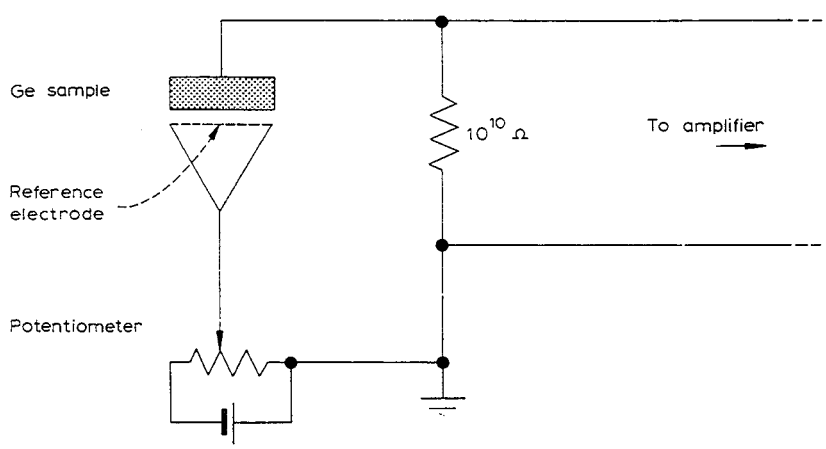


Fig. 9. Schematic diagram of apparatus used by Brattain to measure contact potential and change of contact potential with light.

diffusing to the interior. It was found[13] that the change is usually much larger and often of the opposite sign than predicted by Frenkel's theory, which did not take into account a surface barrier.

Some rather difficult experiments which at the time gave negative results have been carried out successfully much later by improved techniques, as will be described by Dr. Brattain in his talk.

Apparatus used by Brattain to measure contact potential and change in contact potential with illumination is shown in Fig. 9. The reference electrode, generally platinum, is in the form of a screen so that light can pass through it. By vibrating the electrode, the contact potential itself can be measured by the Kelvin method. If light chopped at an appropriate frequency falls on the surface and the electrode is held fixed, the change with illumination can be measured from the alternating voltage developed across the condenser. In the course of the study, Brattain tried several ambient atmospheres and different temperatures. He observed a large effect when a liquid dielectric filled the space between the electrode and semiconductor surface. He and Gibney then introduced electrolytes, and observed effects attributed to large changes in the surface barrier with voltage applied across the electrolyte. Evidently ions piling up at the surface created a very large field which penetrated through the surface states.

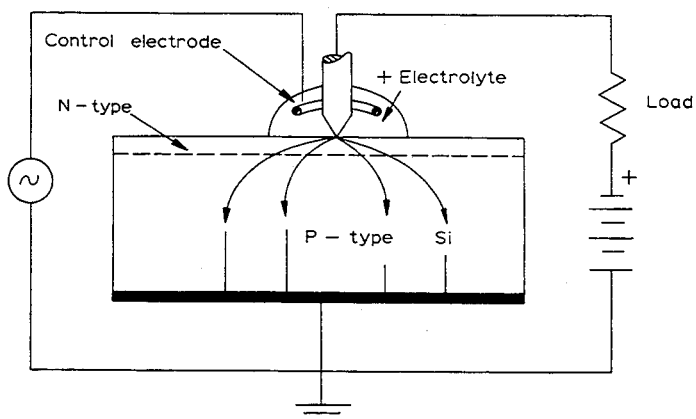


Fig. 10. Diagram of experiment used to observe effect of the field produced by an electrolyte on an inversion layer of n-type conductance at the surface of a p-type silicon block. Negative potential applied to the probe in the electrolyte decreases the number of electrons in the inversion layer and thus the current of electrons flowing to the point contact biased in the reverse direction. Arrows indicate the conventional direction of current flow; electrons move in the opposite direction.

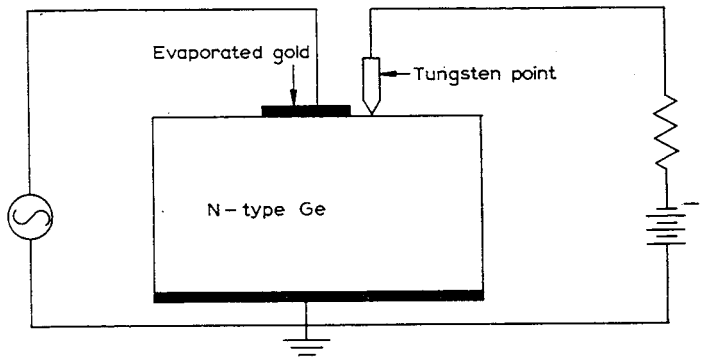


Fig. 11. Diagram of experiment in which the transistor effect was first observed. Positive voltage applied to the gold spot introduced holes into the n-type germanium block which flowed to the point contact biased in the reverse direction. It was found that an increase in positive voltage increased the reverse current. When connected across a high impedance, the change in voltage of the point contact was larger than the change at the gold spot, both measured relative to the base electrode.

Experiments on inversion layers

Use of an electrolyte provided a method for changing the surface barrier, so that it should be possible to observe a field effect in a suitable arrangement. We did not want to use an evaporated film because of the poor structure and low mobility. With the techniques available at the time, it would have been difficult to prepare a slab from bulk material sufficiently thin to observe a sizable effect. It was suggested that one could get the effect of a thin film in bulk material by observing directly the flow in an inversion layer of opposite conductivity type near the surface. Earlier work of Ohl and Scaff indicated that one could get an inversion layer of n-type conductivity on p-type silicon by suitably oxidizing the surface. If a point contact is made which rectifies to the p-type base, it would be expected to make low resistance contact to the inversion layer.

The arrangement which Brattain and I used in the initial tests is shown in Fig. 10. The point contact was surrounded by, but insulated from, a drop of electrolyte. An electrode in the electrolyte could be used to apply a strong field at the semiconductor surface in the vicinity of the contact. The reverse, or high resistance direction is that in which point is positive relative to the block. Part of the reverse current consists of electrons flowing through the n-type inversion layer to the contact. It was found that the magnitude of

this current could be changed by applying a voltage on the electrolyte probe, and thus, by the field effect, changing the conductance of the inversion layer. Since under static conditions only a very small current flowed through the electrolyte, the set-up could be used as an amplifier. In the initial tests, current and power amplification, but not voltage amplification, was observed. As predicted from the expected decrease in number of electrons in the inversion layer, a negative voltage applied to the probe was found to decrease the current flowing in the reverse direction to the contact.

It was next decided to try a similar arrangement with a block of n-type germanium. Although we had no prior knowledge of a p-type inversion layer on the surface, the experiments showed definitely that a large part of the reverse current consisted of holes flowing in an inversion layer near the surface. A positive change in voltage on the probe decreased the reverse current. Considerable voltage as well as current and power amplification was observed.

Because of the long time constants of the electrolyte used, amplification was obtained only at very low frequencies. We next tried to replace the electrolyte by a metal control electrode insulated from the surface by either a thin oxide layer or by a rectifying contact. A surface was prepared by Gibney by anodizing the surface and then evaporating several gold spots on it. Although none made the desired high resistance contact to the block, we decided to see what effects would be obtained. A point contact was placed very close to one of the spots and biased in the reverse direction (see Fig. 11). A small effect on the reverse current was observed when the spot was biased positively, but of *opposite* direction to that observed with the electrolyte. An increase in positive bias *increased* rather than decreased the reverse current to the point contact. The effect was large enough to give some voltage, but no power amplification. This experiment suggested that holes were flowing into the germanium surface from the gold spot, and that the holes introduced in this way flowed into the point contact to enhance the reverse current. This was the first indication of the transistor effect.

It was estimated that power amplification could be obtained if the metal contacts were spaced at distances of the order of 0.005 cm. In the first attempt, which was successful, contacts were made by evaporating gold on a wedge, and then separating the gold at the point of the wedge with a razor blade to make two closely spaced contacts. After further experimentation, it appeared that the easiest way to achieve the desired close separation was to use two appropriately shaped point contacts placed very close together.

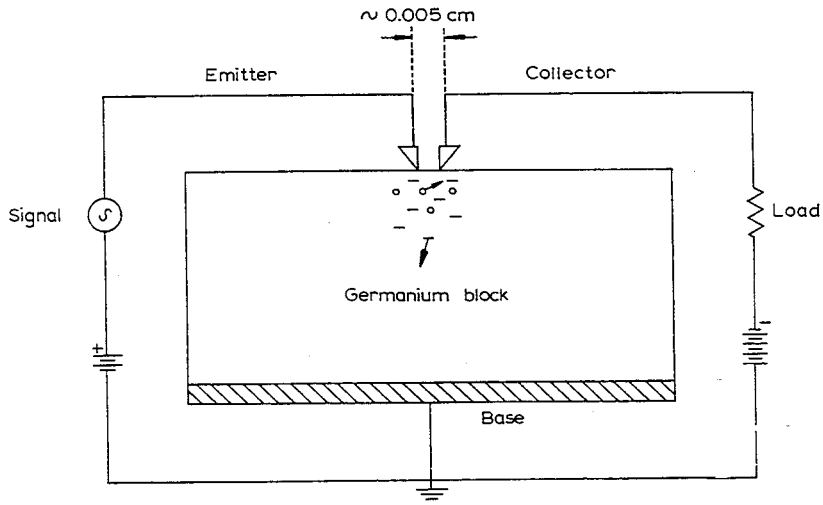


Fig. 12. Schematic diagram of point-contact transistor.

Success was achieved in the first trials; the point-contact transistor was born[14].

It was evident from the experiments that a large part of both the forward and reverse currents from a germanium point contact is carried by minority carriers, in this case holes. If this fact had been recognized earlier, the transistor might have come sooner.

Operation of a point-contact transistor is illustrated in Fig. 12. When operated as an amplifier, one contact, the emitter, is biased with a d.c. voltage in the forward direction, the second, the collector, in the negative or high resistance direction. A third contact, the base electrode, makes a low resistance contact to the block. A large part of the forward current consists of holes flowing into the block. Current from the collector consists in part of electrons flowing from the contact and in part of holes flowing toward the contact. The collector current produces an electric field in the block which is in such a direction as to attract holes introduced at the emitter. A large part of the emitter current, introduced at low impedance, flows in the collector circuit. Biased in the reverse direction, the collector has high impedance and can be matched to a high impedance load. There is thus a large voltage amplification of an input signal. It is found[14] that there is some current amplification as well, giving an overall power gain of 20 db. or more. An increase in hole current at the collector affects the barrier there in such a way as to enhance the current of electrons flowing from the contact.

The collector current must be sufficiently large to provide an electric field to attract the holes from the emitter. The optimum impendance of the collector is considerably less than that of a good germanium diode in the reverse direction. In the first experiments, it was attempted to achieve this by treating the surface so as to produce a large inversion layer of p-type conductivity on the surface. In this case, a large fraction of the hole current may flow in the inversion layer. Later, it was found that better results could be obtained by electrically forming the collector by passing large current pulses through it. In this case the surface treatment is less critical, and most of the emitter current flows through the bulk.

Studies of the nature of the forward and reverse currents to a point contact to germanium were made by making probe measurements of the variation of potential in the vicinity of the contact[15]. These measurements showed a large increase in conductivity when the contact was biased in the forward direction and in some cases evidence for a conducting inversion layer near the surface when biased in the reverse direction.

Before it was established whether the useful emitter current was confined to an inversion layer or could flow through the bulk, Shockley[16] proposed a radically different design for a transistor based on the latter possibility. This is the junction transistor design in which added minority carriers from the emitter diffuse through a thin base layer to the collector. Independently of this suggestion, Shive[17] made a point-contact transistor in which the emitter and collector were on opposite faces of a thin slab of germanium. This showed definitely that injected minority carriers could flow for small distances through bulk material. While transistors can be made to operate either way, designs which make use of flow through bulk material have been most successful. Junction transistors have superseded point-contact transistors for most applications.

Following the discovery of the transistor effect, a large part of research at the Bell Laboratories was devoted to a study of flow on injected minority carriers in bulk material. Much of this research was instigated by Shockley, and will be described by him in the following talk.

Research on surface properties of germanium and silicon, suspended for some time after 1948 because of the pressure of other work, was resumed later on by Brattain and others, and is now a flourishing field of activity with implications to a number of scientific fields other than semiconductors such as adsorption, catalysis, and photoconductivity. This research program will be described by Dr. Brattain in his talk.

It is evident that many years of research by a great many people, both before and after the discovery of the transistor effect, has been required to bring our knowledge of semiconductors to its present development. We were fortunate enough to be involved at a particularly opportune time and to add another small step in the control of Nature for the benefit of mankind. In addition to my colleagues and to others mentioned in the lecture, I would like to express my deep gratitude to Drs. M. J. Kelly and Ralph Bown for the inspired leadership of the Laboratories when this work was done.

1. A. H. Wilson, *Proc. Roy. Soc. London*, A 133 (1931) 458; A 134 (1931) 277; A 136 (1932) 487.
2. J. Frenkel, *Physik. Z. Sowjetunion*, 8 (1935) 185.
3. N. F. Mott, *Proc. Roy. Soc. London*, A 171 (1939) 27.
4. W. Schottky, *Z. Physik*, 113 (1939) 367; 118 (1942) 539.
5. B. Davydov, *J. Tech. Phys. U.S.S.R.*, 5 (1938) 87.
6. R. Hilsch and R. W. Pohl, *Z. Physik*, 111 (1938) 399.
7. Amplifiers based on the field-effect principle had been suggested earlier in the patent literature (R. Lillienfeld and others), but apparently were not successful. Shockley's contribution was to show that it should be possible according to existing semiconductor theory to make such a device. An early successful experiment is that of W. Shockley and G. L. Pearson, *Phys. Rev.*, 74 (1948) 232.
8. J. Bardeen, *Phys. Rev.*, 71 (1947) 717.
9. A review is given in the lecture of Dr. Brattain, this volume, p. 337.
10. G. L. Pearson and J. Bardeen, *Phys. Rev.*, 75 (1949) 865.
11. *See* K. Lark-Horovitz, *Elec. Eng.*, 68 (1949) 1047.
12. W. H. Brattain and W. Shockley, *Phys. Rev.*, 72 (1947) 345.
13. W. H. Brattain, *Phys. Rev.*, 71 (1947) 345.
14. J. Bardeen and W. H. Brattain, *Phys. Rev.*, 74 (1948) 230; 75 (1949) 1208.
15. W. H. Brattain and J. Bardeen, *Phys. Rev.*, 74 (1948) 231.
16. W. Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand Co., Inc., New York, 1950, p. 86.
17. J. N. Shive, *Phys. Rev.*, 75 (1949) 689.

Biography

John Bardeen was born in Madison, Wisconsin, on May 23, 1908, son of Dr. Charles R. Bardeen, and Althea Harmer. Dr. Bardeen was Professor of Anatomy, and Dean of the Medical School of the University of Wisconsin at Madison. After the death of Althea, when John was about twelve years old, Dr. Bardeen married Ruth Hames, now Mrs. Kenelm McCauley, of Milwaukee, Wisconsin.

John Bardeen attended the University High School at Madison for several years, but graduated from Madison Central High School in 1923. This was followed by a course in electrical engineering at the University of Wisconsin, in which much extra work was taken in mathematics and physics. After being out for a term while working in the engineering department of the Western Electric Company at Chicago, he graduated with a B.S. in Electrical Engineering in 1928. He continued on at Wisconsin as a graduate research assistant in electrical engineering for two years, working on mathematical problems in applied geophysics and on radiation from antennas. It was during this period that he got his first introduction to quantum theory from Professor J. H. Van Vleck.

Professor Leo J. Peters, under whom the research in geophysics was done, took a position at the Gulf Research Laboratories in Pittsburgh, Pennsylvania, and Bardeen followed him there and worked during the next three years (1930–1933) on the development of methods for the interpretation of magnetic and gravitational surveys. This was a stimulating period in which geophysical methods were first being applied to prospecting for oil.

Because he felt his interests were more in pure than in applied science, Bardeen resigned his position at Gulf in 1933 to take graduate work in mathematical physics at Princeton University. It was here under the leadership of Professor E. P. Wigner, that he first became interested in solid state physics. Before completing his thesis (on the theory of the work function of metals) he was offered a position as Junior Fellow of the Society of Fellows at Harvard University. He spent there the next three years, 1935–1938, working with Professors Van Vleck and Bridgman on problems in cohesion and

electrical conduction in metals, and also did some work on level density of nuclei. The Ph.D. degree at Princeton was awarded in 1936.

From 1938–1941, Bardeen was an Assistant Professor of Physics at the University of Minnesota and from 1941–1945 a civilian physicist at the Naval Ordnance Laboratory in Washington, D.C. Work done during the war was on influence fields of ships for application to underwater ordnance and mine-sweeping. After the war, in late 1945, he joined the solid state research group at the Bell Telephone Laboratories, and remained there until 1951, when he was appointed Professor of Electrical Engineering and of Physics at the University of Illinois. Since 1959 he has also been a member of the Center for Advanced Study of the University.

Main fields of research since 1945 have been electrical conduction in semiconductors and metals, surface properties of semiconductors, theory of superconductivity, and diffusion of atoms in solids. In 1957, Bardeen and two colleagues, L. N. Cooper and J. R. Schrieffer, proposed the first successful explanation of superconductivity. Much of his research effort since that time has been devoted to further extensions and applications of the theory.

He is a Fellow of the American Physical Society, has been (1954–1957) a member of its Council, and on the Editorial Board of *The Physical Review* and *Reviews of Modern Physics*. From 1959–1962, he served as a member of the United States President's Science Advisory Committee.

Bardeen was elected to the National Academy of Sciences in 1954. Honours include the Stuart Ballentine Medal of the Franklin Institute, Philadelphia (1952) and the John Scott Medal of the City of Philadelphia (1955), both awarded jointly with Dr. W. H. Brattain, the Buckley Prize of the American Physical Society (1955) and D.Sc. (Hon.) from Union College and from the University of Wisconsin. He received the Fritz London Award for work in low temperature physics in 1962.

Bardeen married Jane Maxwell in 1938. They have three children, James Maxwell, William Allen and Elizabeth Ann.

WILLIAM SHOCKLEY

Transistor technology evokes new physics

Nobel Lecture, December 11, 1956

Introduction

The objective of producing useful devices has strongly influenced the choice of the research projects with which I have been associated. It is frequently said that having a more-or-less specific practical goal in mind will degrade the quality of research. I do not believe that this is necessarily the case and to make my point in this lecture I have chosen my examples of the new physics of semiconductors from research projects which were very definitely motivated by practical considerations.

An important fraction of United States industry adheres to the idea that research of a fundamental character is worthwhile from a practical point of view. This is outstandingly the case at Bell Telephone Laboratories where my co-prizewinners and I, together with our many colleagues, carried out the work described in these lectures. The attitude of Bell Telephone Laboratories has undoubtedly resulted to a substantial degree from the viewpoints of the four men who have been its research directors and subsequently its presidents. Each of these men, H. D. Arnold, F. B. Jewett, O. E. Buckley, and M. J. Kelly, has also been active and effective in governmental or civic affairs. All had obtained a thorough indoctrination in the research viewpoint in the course of their doctorate training in physics. My personal contact with two of these men had a significant influence on my planning of semiconductor research programs as I shall mention below.

My decision to come to Bell Telephone Laboratories immediately after obtaining my Ph.D. in 1936 was strongly influenced by the fact that my supervisor would be C. J. Davisson. Upon my arrival I was assigned by Dr. M. J. Kelly to an indoctrination program in vacuum tubes. In the course of this program Dr. Kelly spoke to me of his ideal of doing all telephone switching electronically instead of with metal contacts. Although I did not choose to continue work on vacuum tubes and was given freedom to pursue basic research problems in solid-state physics, Dr. Kelly's discussion left me continually alert for possible applications of solid-state effects in telephone

switching problems. Insofar as my contribution to transistor electronics has hastened the day of a fully electronic telephone exchange, it was strongly stimulated by the experiences given me during my early years at the Laboratories.

Before leaving the subject of research in industry, I would like to express some viewpoints about words often used to classify types of research in physics; for example, pure, applied, unrestricted, fundamental, basic, academic, industrial, practical, etc. It seems to me that all too frequently some of these words are used in a derogatory sense, on the one hand to belittle the practical objectives of producing something useful and, on the other hand, to brush off the possible long-range value of explorations into new areas where a useful outcome cannot be foreseen. Frequently, I have been asked if an experiment I have planned is pure or applied research; to me it is more important to know if the experiment will yield new and probably enduring knowledge about nature. If it is likely to yield such knowledge, it is, in my opinion, good fundamental research; and this is much more important than whether the motivation is purely esthetic satisfaction on the part of the experimenter on the one hand or the improvement of the stability of a high-power transistor on the other. It will take both types to « confer the greatest benefit on mankind » sought for in Nobel's will.

The five basic imperfections

Before discussing the selected group of experiments in transistor physics, I shall extend somewhat farther Dr. Bardeen's remarks[1] about the characteristics and interactions of electrons, holes, donors and acceptors. For this purpose I shall use a reference to water solutions of acids, bases and salts as an aid to exposition.

The dissociation of pure water to positive hydrogen ions and negative hydroxyl ions satisfies the mass-action law

$$[H^+]\,[OH^-] = \text{function of } T$$

where the concentration of H_2O is so nearly constant for dilute solutions that it may be taken as constant and not shown explicitly. This equation has as its analogue

$$n \cdot p = f(T)$$

for a semiconductor where n and p are the electron and hole concentrations. The equation is accurate, provided neither n nor p is so large that its statistics become degenerate. The normal electron-pair bond here plays the role of an undissociated water molecule. In pure deionized water and in a pure semiconductor, electrical neutrality requires that the concentration of the positive charges equals that of the negative charges:

$$[H^+] = [OH^-] \quad \text{and} \quad p = n$$

A semiconductor which is so pure that impurities do not affect the property being considered is called *intrinsic* in regard to such a property. Using the subscript i to designate this case, we have

$$p_i = n_i \quad \text{and} \quad n_i^2 = f(T)$$

The concentration n_i is referred to as the intrinsic concentration.

The chemical analogue to an n-type semiconductor is a base and the charge neutrality equations are

$$[H^+] + [Na^+] = [OH^-]$$
$$p + N_d = n$$

where N_d is the concentration of donors and it is assumed that all donors are fully ionized.

Similarly a p-type semiconductor is analogous to an acid:

$$[H^+] = [OH^-] + [Cl^-]$$
$$p = n + N_a$$

where N_a is the acceptor concentration.

A neutral salt also has its analogue which is called a *compensated* semiconductor. For this case the donor and acceptor concentrations are equal so that the equation for electrical neutrality

$$p + N_d = n + N_a$$

reduces to equality of n and p so that each is equal to n_i. The electrical conductivity of a perfectly compensated silicon or germanium crystal is almost equal to that of an intrinsic crystal; there may be a small difference due to the reduction of mobility by scattering by the charged ions. The difference between the low conductivity of the perfectly compensated semiconductor and the high conductivity of a neutral salt solution arises from the fact that the donors and acceptors are tied in place in a semiconductor while the cations and anions of a salt have mobilities comparable to that of an OH-ion.

Compensation, or rather overcompensation, plays a vital role in the manufacture of semiconductor devices. For example, it permits the conversion of n-type material to p-type by the addition of acceptors without the prior removal of donors. A crystal grown from an n-type melt may be changed to p-type by adding an excess of acceptors.

The words *majority* and *minority* are frequently useful in discussing semiconductors. In an n-type specimen, the majority carriers are electrons and the minority carriers are holes.

Holes, electrons, donors and acceptors represent four of the five classes of imperfections that must be considered in semiconductor crystals in order to understand semiconductor effects. The fifth imperfection has been given the name *deathnium*. The chemical analogue of deathnium is a catalyst. In the case of water as the analogue to a crystal, there is, so far as I know, no important corresponding catalyst. What deathnium does is to hasten the establishment of equilibrium between holes and electrons. If, due to the various possible disturbances important in transistor electronics, the concentration of minority carriers is, for example, substantially raised, then the minority carriers will combine with the majority carriers to produce normal electron-pair bonds, by this means restoring equilibrium. Deathnium catalyzes this recombination process. The symbols for the five imperfections are shown in Table 1.

Table 1.

1. $-$ (excess) electron
2. $+$ hole
3. deathnium
4. $\oplus$ donor
5. $\ominus$ acceptor

The role of deathnium can be illustrated in terms of the phenomenon of photoconductivity. If light shines on a germanium crystal, then the pairs of

electrons and holes that are formed will impart a conductivity to the crystal. This conductivity is known as photoconductivity. If the source of light is removed, the photoconductivity will die away, due to the recombination of the holes and the electrons. Thus, if an electron falls into an incomplete bond, one hole-electron pair will be eliminated.

The photoconductivity dies away with a characteristic time known as the lifetime. Thus, after the light is turned off, the photoconductivity will drop to approximately one-half its value in one lifetime. This process continues with a reduction of approximately one-half in each subsequent period of one lifetime.

If the process of recombination of holes and electrons were a direct one, the lifetime would be the same in all germanium crystals. It is found experimentally, however, that two otherwise very similar germanium crystals will have lifetimes that differ by as much as 1,000 fold. In one crystal, the lifetime may be a millisecond, whereas in another it may be a microsecond. This variation in lifetime requires the presence of some sort of imperfection which catalyzes the recombination of the holes and the electrons.

Actually, there are several forms of deathnium. For example, if electrons having an energy of several million electron volts fall upon a germanium crystal, the lifetime is subsequently reduced[2]. From the investigation at Purdue University, it is known that such bombardment produces disorder of the germanium atoms[3]. A high-energy electron can eject a germanium atom bodily from its normal position in the crystal structure, thus leaving a vacancy behind, where there should be an atom, and causing the ejected atom to become either an extra atom or an interstitial atom fitting into a place in the structure which would normally be empty. It has been found at Bell Telephone Laboratories that these disordering effects function as deathnium. It has also been found that copper and nickel chemical impurities in the germanium produce marked reductions in lifetime[4].

The way in which deathnium catalyzes the recombination process is in-

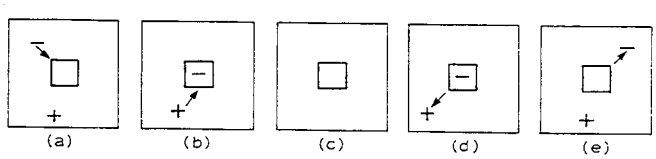


Fig. 1. A recombination center (deathnium) captures alternately an electron and a hole and thus catalyzes their recombination, as shown in parts (*a*), (*b*), and (*c*). The thermally activated generation process is shown in (*d*) and (*e*).

dicated in Fig. 1. In part (*b*) of this figure, an electron is captured by a deathnium center. The deathnium center thus becomes a baited trap which is ready to capture a hole. If a hole comes near to the deathnium center, the electron can drop into it, thus forming a normal covalent bond, and the deathnium center is then uncharged and ready to repeat the process.

It is characteristic of all microscopic processes that they may go backwards as well as forwards. Thus, the deathnium center may generate hole-electron pairs as well as eliminate them. The generation process is indicated in parts (*d*) and (*e*) of Fig. 1. In part (*d*) the deathnium center captures an electron from an adjoining normal electron-pair bond. This produces a hole which wanders off. Somewhat later, the deathnium center ejects the electron and thus reverts to its empty state in which it is ready either to recombine or to generate another hole-electron pair.

Under conditions of thermal equilibrium, both the recombination process and the generation process proceed continuously. The energy required to generate the hole-electron pair is furnished by the thermal energy of vibration of the atoms in the germanium crystal. The condition of thermal equilibrium is achieved when the two processes balance. For germanium at room temperature, this leads to a conductivity of about 0.02 ohm^{-1} cm^{-1}.

Since the concentration of holes and electrons under equilibrium conditions is governed by a mass-action law, the product np is independent of the concentration of deathnium. For example, if the concentration of deathnium is doubled, both the rate of generation and the rate of recombination are doubled, but the equilibrium concentrations of holes and electrons are unaltered.

Evidence that the deathnium mechanism shown in Fig. 1 is correct has been obtained by studying the dependence of the rate of recombination upon hole and electron densities[5]. These studies are found to be in general agreement with the predictions based on the statistical model of Fig. 1.

The field effect

The experiment which played the largest role in stimulating the transistor electronics program at Bell Telephone Laboratories was the so-called field-effect experiment. I devised this experiment as a direct result of trying to invent a semiconductor amplifier having separate input and output circuits. From the then known properties of semiconductors, I concluded that a thin

film of silicon or germanium would have a substantial change in its conductivity if it were made into one of a pair of condenser plates and the condenser strongly charged. The surface charge, if represented as a change in mobile carriers, could appreciably increase or decrease the conductance of the film.

A number of experiments were initially tried using evaporated layers and other thin layers. These all gave negligible effects and progress was at a standstill until Bardeen proposed his theory of surface states to explain the lack of an observable effect.

Bardeen's model also explained a number of other previously mysterious phenomena[6] and led to the suggestion of doing the field-effect experiment at liquid-air temperature to immobilize the surface states. This gave the first positive result. How this led to a sequence of experiments culminating in the point-contact transistor is discussed in Dr. Bardeen's lecture[1]. The first publication of a positive effect was made by G. L. Pearson and the writer in 1948[7].

At the present time the field-effect experiment is playing a very important role in measuring the properties of semiconductor surfaces[8].

On the practical side the field-effect has been utilized to make transistor amplifiers having interesting properties quite different from those of junction transistors.

Injection and drift

At the time of the discovery of the point-contact transistor by Bardeen and Brattain there were a number of unresolved questions regarding its mode of operation. The original transistor showed evidence that the coupling between the input or emitter point and the output or collector point took place through conduction in a thin surface layer of conductivity type opposite to the underlying or base material. Somewhat later the idea that the emitter point might actually be injecting minority carriers into the body of the semiconductor developed. The development of this idea came as a result of two independent events: the invention of the junction-transistor structure by the present writer (as discussed below, injection plays an essential role in the junction transistor) and the observation by J. N. Shive[9] that transistor action could occur with the points on opposite sides of a thin slab of semiconductor.

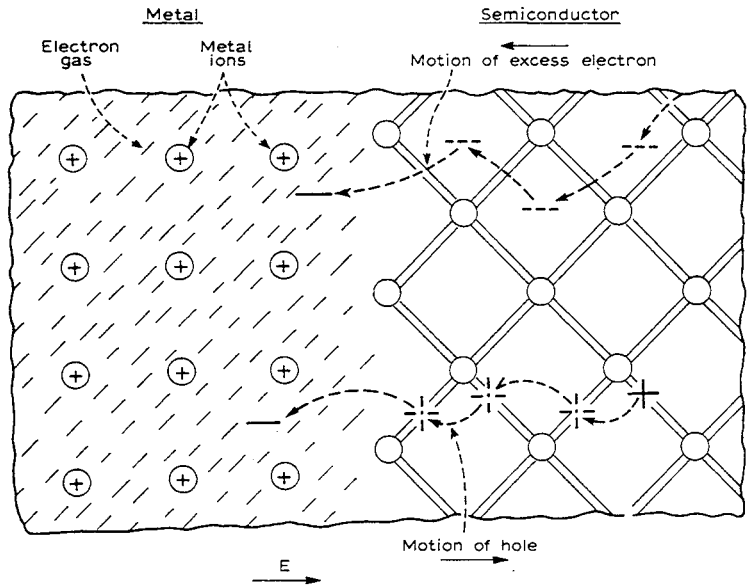


Fig. 2. Two possible mechanisms for current flow near an emitter point as described in text.

In order to test the model of carrier injection, J. R. Haynes and the author collaborated in the drift-mobility experiment or «Haynes' experiment» on germanium specimens[10]. In order to understand the significance of the experiment in elucidating transistor action the mechanism of current flow at the metal-semiconductor contact must be considered.

In Fig. 2, the metal is represented in a highly pictorial fashion. The valence electrons in a metal are thought of as forming an electron gas, which permeates the entire structure. Thus, the electrons are not held in position in valence bonds as they are in an insulator. The electron gas can flow freely through the structure of the metal, and this fact accounts for the high conductivity of metals. In the upper part of Fig. 2 one of the processes for removing electrons from the semiconductor is represented. Since the semiconductor is n-type, it contains excess electrons; these excess electrons may be drawn to the metal by its positive charge and thus enter the metal to produce a current of electrons flowing out of the emitter point through the connecting lead.

Another possible mechanism for electron transfer from semiconductor to metal is shown in the lower part of Fig. 2. In this case, an electron is withdrawn from one of the valence bonds adjacent to the metal. This process

also transfers an electron from the semiconductor to the metal, but when the transfer occurs, a hole is left behind. The hole is repelled by the positive charge on the emitter contact and moves deeper into the semiconductor.

Both of the processes discussed above have the same effect so far as the metal emitter point and conducting leads to the emitter point are concerned. Both produce a net flow of electrons from semiconductor to the emitter point and through the leads to the emitter terminal. It is thus evident that some more subtle experiment than simply measuring the current to the emitter point is necessary to show that both processes of electron removal from the semiconductor occur. Suitable experiments have been planned and performed, with the result that it is possible to show that both of the processes of Fig. 2 occur and also to determine the fraction of current carried by each. In fact, in a good emitter point it can be shown that more than 90 per cent of the current is carried by the process which injects holes into the semiconductor, and less than 10 per cent by the process which removes electrons.

In an ideal emitter point, all of the current would be carried by the hole-injection process. The reason for this result is that the electron-removal process does not disturb the state of affairs within the semiconductor. If electrons are removed from the semiconductor in the neighborhood of the emitter point, they are promptly replaced by electrons flowing from more distant parts of the semiconductor, and these electrons in turn are replaced by other electrons flowing in from whatever contact to the semiconductor completes the electrical-current path or circuit. In the hole-injection process the situation is quite different. Normally, the number of holes in the n-type semiconductor is negligible so far as a contribution to the conductivity is concerned. However, when electrons are removed from the valence bonds and holes are injected, relatively large numbers of holes will be introduced. The conductivity of the semiconductor will be increased in the neighborhood of the emitter point in much the same fashion that it would be if light were to shine on the semiconductor and produce hole-electron pairs. This disturbance in the electronic structure can be used to produce amplifying action in the transistor.

Instead of discussing the quantitative experiment which is used to distinguish between the two processes shown in Fig. 2, I shall describe a qualitative experiment which shows that hole injection does occur at an emitter point. This experiment permits quantitative studies to be made of the behavior of holes and provides a method for the direct measurement of diffusion and drift.

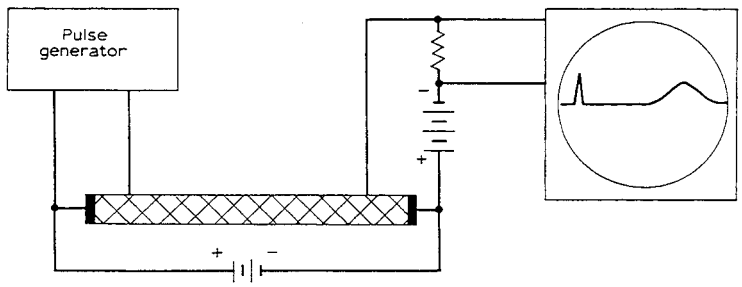


Fig. 3. Schematic representation of experiment to observe the drift and diffusion of injected holes in n-type germanium.

The experimental arrangement, which was first carried out in this form by J. R. Haynes, is illustrated diagrammatically in Fig. 3. The germanium specimen is in the form of an elongated point-contact transistor. There is, however, an extra contact on the base. The germanium is present as a rod, about 1/32 of an inch in cross section and approximately 1 inch long. A « sweeping field » is applied from end to end of the rod by a battery. This field acts in such a direction as to draw electrons from right to left through the rod. If any holes were introduced in the rod, they would drift from left to right.

When the pulse generator at the left-hand point contact, or emitter point, operates, the emitter point is biased positive and thus in the forward direction. According to the ideas presented in Fig. 2, this condition causes holes to be injected into the rod. These holes are then drawn down the rod by the sweeping field. After a time they arrive in the neighborhood of the collector point, which, as the figure shows, is biased negative. It thus tends to attract holes, and some of the holes flow to the collector point and thus contribute to the current flowing in the collector circuit. This current flows through a resistor, and the voltage across the resistor is applied to the vertical plates of a cathode-ray oscilloscope.

Under operating conditions, the operation of the pulse generator is accomplished electronically and is synchronized with the functioning of the oscilloscope, so that just before the switch is closed, the electron beam in the oscilloscope starts to move across the tube face from left to right. At time t_1 the switch to the emitter point is closed for a brief moment; the time of closing is indicated by a « pick up » signal on the face of the oscilloscope. After this nothing happens until time t_2 when some of the holes arrive at the collector point; the concentration of holes builds up for a moment and

then decays as the group of holes injected at time t_2 passes by the collector point. The arrival pulse at the collector point is not so sharp as the « pick up » pulse because the holes, which were injected approximately at one point and at the same time, spread out by diffusion so that by the time the group of holes reaches the collector point it is relatively large in extent along the rod.

It is evident that this experiment permits observation and measurement of both diffusion and drift. It is possible to measure the distance between the points and the electric field between the points; by calibrating the oscilloscope, the time of travel may be measured. Thus the drift velocity may be measured directly, verifying the fact that the disturbance occurring at the emitter point behaves precisely as would be expected if the emitter point injected small numbers of positive carriers into the rod. For example, if the distance between the points is doubled, the time lag between pick-up at t_1 and the arrival of the pulse is also doubled. This result shows that the carriers drift at a constant drift speed in the rod. But if the sweeping field is doubled, the time lag is cut in half. This fact shows that the speed of the carriers is proportional to the electric field. If the polarity of the sweeping field is reversed, we would expect the injected carriers to be drawn to the left in the filament so that none arrive at the collector point, and it is found experimentally that this is true.

As was indicated above, the spread of the time of arrival of holes is a measure of the diffusion constant. From studies of the dependence of this spread upon the transit time from emitter to collector, it can be verified that the holes spread out in accordance with the laws expected for diffusion phenomena. The value of the diffusion constant D can also be measured.

J. R. Haynes and his colleagues have performed various experiments of this sort. They have also experimented with the case of electron injection into p-type germanium and have dealt with the two corresponding cases for silicon. The values of mobility and diffusion constant which they obtain in this way[11] are tabulated in Table 2.

Table 2. Mobilities in cm^2/volt sec and diffusion constants in cm^2/sec.

	Electrons		*Holes*	
	μ	D	μ	D
Silicon	1.200	30	250	6.5
Germanium	3.600	93	1.700	43

It should be noted from Table 2 that in each case the ratio of diffusion constant to mobility is approximately 1/40, and the dimensions of this quantity are in volts. In other words, the ratio of D to μ is 25 millivolts. This value has a fundamental significance, and the relationship between D and μ is commonly known as the Einstein relationship. This relationship has recently been investigated in detail, by the means described above, for germanium[12]. The significance of this value of 25 millivolts is that an electron moving with random thermal energy will, on the average, be just capable of surmounting a potential hill of 25 millivolts. In other words, 25 millivolts is the electrostatic potential corresponding to thermal energy for one electron. Put in another way, it can be stated that if any electron was set in motion with thermal energy in free space against any electric field, the electron would be slowed down by the electric field and by the time it had moved 25 millivolts against the field its velocity would be brought to zero and it would start to move in the opposite direction. The fact that a value of 25 millivolts is obtained shows that the charge of the carriers which are drifting and diffusing in the Haynes experiment is the electronic charge. If it were half or twice this value, for example, the ratio of D to μ would be 50 or 12.5 millivolts, respectively.

When the Haynes experiment was first carried out, the procedure was varied by placing the two points on opposite sides of the filament and by using filaments of much greater width than thickness. All these experiments indicated that the interaction between the points took place through the bulk rather than over the surface.

Hot electrons and Ohm's law

Another fundamental experiment on the behavior of electrons and holes in semiconductors had its origin in a practical aim. Sometime during 1948 in an attempt to see how semiconductors could be made to amplify at very high frequencies, I came upon an idea which was independently discovered and first published as the *Staueffekt* by H. A. Krömer[13]. Krömer does not appear to have considered the effect from the point of view of amplification, and its utilization for this purpose was first published in 1954[14].

The basic principle of the staueffect may be understood as follows: a hole in the valence band will lose energy to phonons at a maximum rate $P(\max)$

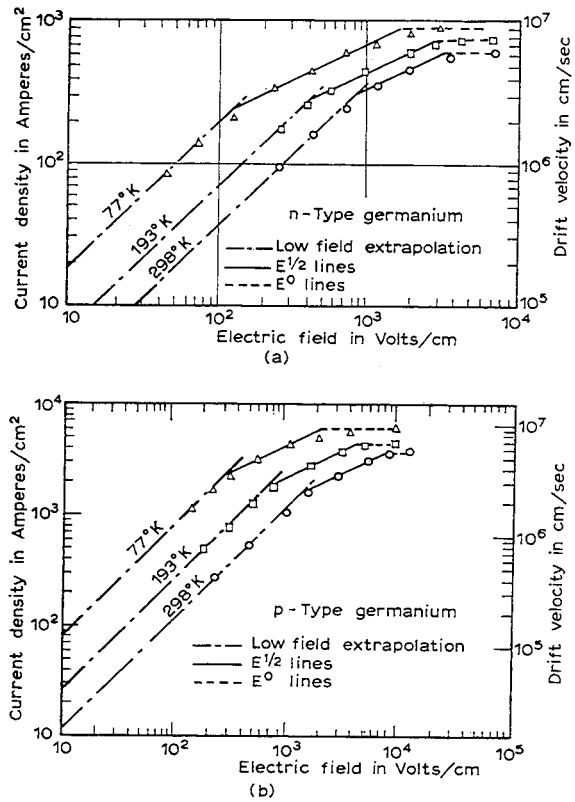


Fig. 4. (*a*, *b*). Current density as a function of electric field. (*a*) n-type germanium; (*b*) p-type germanium.

when its energy is somewhere near the middle of the band. If subjected to an electric field E, its drift velocity v_d must be limited by

$$q E v_d < P(\max)$$

since it cannot gain energy indefinitely. What happens in more detail is this: if the hole gains too much energy, it moves to such low states in the valence band that its effective mass changes sign and it is decelerated instead of accelerated by the field.

The negative resistance effects were not observed. However, experimental situations were produced by E. J. Ryder and W. Shockley[15] in which the random energies of holes and electrons were raised to equivalent temperatures of many thousands of degrees.

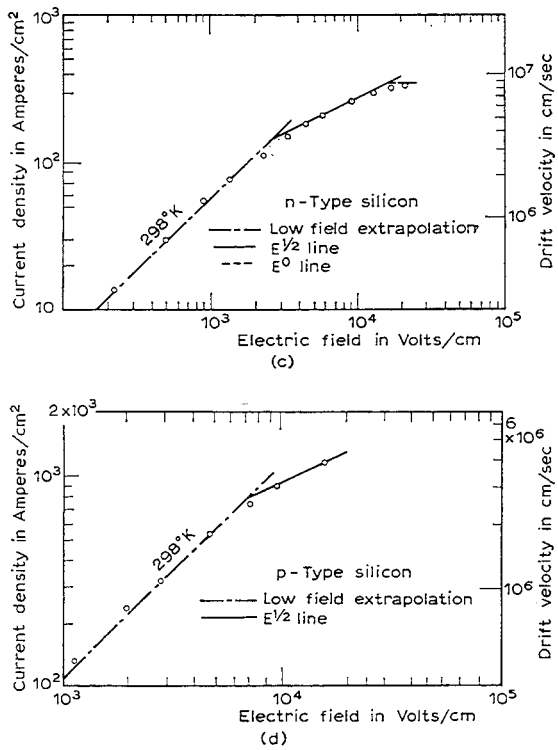


Fig. 4. (*c*, *d*). Current density as a function of electric field.
(*c*) n-type silicon; (*d*) p-type silicon.

These effects represent a fundamental deviation from Ohm's law in the sense that they result solely from the magnitude of the electric field and not from side effects such as changing numbers of holes and electrons due to changing temperature.

Fig. 4 shows several curves representing how the drift velocity of holes and electrons varies with electric field as reported by Ryder[16]. Although the general features of the curves are well understood, they are not yet adequately explained quantitatively[17]. The discrepancy between simple theory and experiment probably results from the complex structure of the energy surfaces for holes and electrons, that is, from deviations from the single-effective-mass model of semiconductors.

From a practical point of view, non-linearities in the drift velocity are important in calculating the characteristics of semiconductor devices.

The p-n junction and single crystals

The rectifying contacts of chief importance in the early days of transistor physics were made by pressing pointed wires on the semiconductor and subsequently conditioning them («forming») by passing heavy current. It was difficult at the time to interpret these procedures from an atomic point of view and this is still true at present.

An important decision in the year following the invention of the transistor was to place emphasis upon p-n junctions in the interior of the semiconductor rather than upon point contacts. From earlier work it was known that such junctions were rectifying and photosensitive. However, their characteristics were not entirely understood in terms of atomic models. Since a p-n junction can occur in a single crystal simply by a change in impurity content from donor dominated in the n-region to acceptor dominated in the p-region, it is a far simpler structure than a metal semiconductor contact and its behavior highly predictable on theoretical grounds. It seemed logical, therefore, to attempt to understand p-n junctions first and point contacts later.

Another reason for choosing to emphasize p-n junctions was the possibility of producing junction transistors, a possibility published in 1949[18].

A p-n junction is the simplest so-called *compositional* structure in semiconductor electronics. By a compositional structure is meant a single crystal of semiconductor in which the composition in terms of the donor and acceptor content varies in a systematic and controlled way. Before describing the theory of the functioning of a p-n junction, I would like to say something about the way in which p-n junctions whose behavior was satisfactory from a theoretical point of view were first achieved at Bell Telephone Laboratories. This history also is an example of the interaction of practical needs on a research program.

To begin, with attempts were made, particularly by M. Sparks, to produce p-n junctions by allowing molten drops of germanium of one conductivity type to fall upon a heated block of germanium of the other conductivity type. Although p-n junctions were obtained by these means, their characteristics failed to live up to the predictions of theory. (The problems were almost certainly those of cleanliness – the importance of copper was not known at the time.)

As a result of the needs of the Development Department for uniform material for the purpose of making experimental transistors, a program was

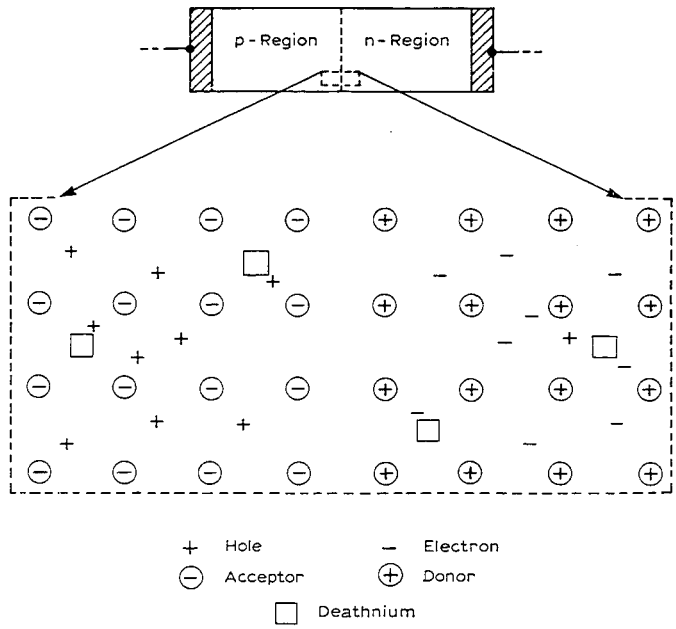


Fig. 5. A p-n junction and the distribution of imperfections in it. (For simplicity, compensating pairs of donors and acceptors are not shown.)

set up to grow large single crystals of germanium. It is interesting to note that the actual decision and emphasis on this program was due principally to J. A. Morton, who headed the transistor-development group, rather than by my own group or other research groups. I felt at the time that scientific studies could adequately be carried out by cutting small specimens from polycrystalline masses of material. As a result of Morton's urging, G. K. Teal, with the collaboration of J. B. Little, made an experimental apparatus for « pulling » single crystals of germanium from a melt of germanium heated in a graphite crucible into which was dipped a small single-crystal seed[19]. The recent advances of transistor science and technology are founded on these crystals. The addition of acceptors to an n-type melt during pulling changed the solidifying material from n- to p-type and gave the first good p-n junctions.

Another extremely important development in the preparation of materials should be mentioned. This is the method of *zone refining* invented by W. G. Pfann, also of Bell Telephone Laboratories[20]. Noting that impurities tend to be more soluble in molten germanium than in solid germanium.

Pfann devised a system of repetitive purification by crystallization. By making an elongated graphite boat and providing means to heat locally a small length or zone of it, he was able to pass molten zones successively from one end to the other of the germanium and by this means to sweep impurities clear of the major portion of the crystal and to concentrate them near one end. By this means germanium crystals having one impurity atom in about 10^{10} germanium atoms have been produced. The density of impurities in these crystals is thus comparable to the density of molecules in a gas at a pressure of 10^{-5} millimeters of mercury. It is appropriate to call zone refining the vacuum pump of transistor electronics.

Fig. 5 represents a p-n junction. In discussing its electrical properties, we will be concerned with the five kinds of imperfections shown in the lower part of the figure. From a mechanical point of view the crystal is practically homogeneous and perfect. A typical concentration for impurities in the crystal might be 10^{15} cm^{-3}. This density of imperfections is so low that if one were to traverse a row of atoms from end to end in the crystal one would, on the average, strike only about ten imperfections. Thus the crystal structure itself is only slightly altered by the presence of the imperfections. From the electrical point of view, on the other hand, the imperfections have profound effects.

As is shown in Fig. 5, the electrons are found chiefly in the n-region where they neutralize the chemical space charge of the donors, and the holes are similarly found in the p-region. In order for this situation to persist, as it does under conditions of thermal equilibrium, there must be an electric field present in the crystal. The idea that an electrical field is present in a conductor under conditions of thermal equilibrium is at first surprising. However, the necessity for such an electric field can readily be seen in terms of Fig. 5. Let us first of all suppose that no electric field is present across the junction; then as a result of diffusion, holes will cross the junction from left to right and electrons will similarly diffuse across the junction from right to left. As a result, there will be a net transfer to the right of positive charge across the junction. This will result in an electric field which will build up to just such a point that further current flow is prevented.

The electric field in the p-n junction is represented in Fig. 6 (a) in terms of an electrostatic potential ψ. The remaining parts of the figure show how this electric field and electrostatic potential arise from the charge densities involved. The chemical charge densities are shown in parts (b) and (c) of the figure. In this example it has been assumed, as represented by the N_d

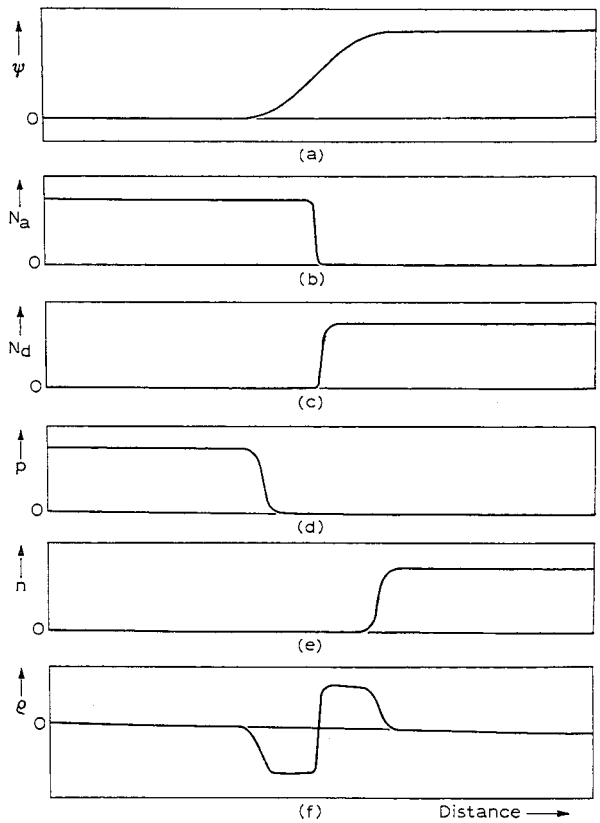


Fig. 6. Potential and charge distribution in a p-n junction.

and N_a curves, that the transition from n-type to p-type occurs abruptly at the junction and that compensation of one impurity type by another is not involved. In the presence of the electrostatic potential shown in part (*a*) of the figure the holes tend to be pushed to the left. As a result, the hole density drops to a small value before the junction itself is reached. Electrons having a negative charge tend to move to the points of highest electrostatic potential, and thus they also are not found near the center of the junction. As a consequence, the chemical charge density is not compensated by holes or electrons in the immediate neighborhood of the junction. Thus an electrostatic dipole layer is formed at the junction, the charge density being as shown in part (*f*) of the figure. This dipole layer is seen to be of just the nature necessary to produce the electrostatic potential shown in part (*a*).

Mathematically, what is done in order to determine the shape of the elec-

trostatic potential in Fig. 6(*a*) is to solve a differential equation. If the dependence of electrostatic potential upon distance is regarded as the unknown, then from it and certain principles of statistical mechanics it is possible to write an expression for the charge density due to the holes and the electrons. These charge densities can be combined with those due to the chemical imperfections in order to obtain a differential equation for the electrostatic potential. This differential equation is Poisson's equation, which relates derivatives of the electrostatic potential to the charge density. When this equation is solved, it is found that the situation in the p-n junction under thermal equilibrium conditions is as represented in Fig. 6.

Under conditions of thermal equilibrium no net current of either holes or electrons will flow across the junction. It is advantageous, however, to consider this equilibrium situation as arising from compensating currents. We shall illustrate this by considering the flow of holes back and forth across the junction. Although the density of holes is small in the n-region, it is still appreciable and plays a vital role in the behavior of the p-n junction. Let us consider the course of a hole that arrives in the n-region by climbing the potential hill as illustrated in Fig. 7. Having climbed the hill and arrived at the plateau of constant electrostatic potential in the n-type region, it will then proceed to move by a random diffusive motion. The most probable outcome of this motion will be that it will diffuse to the steep part of the

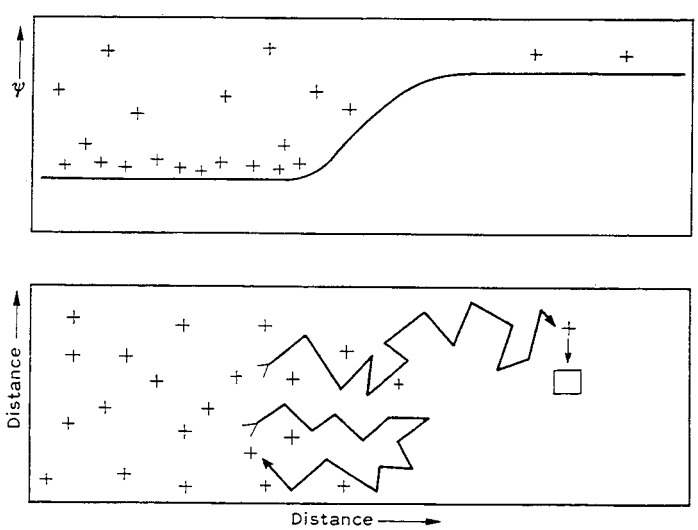


Fig. 7. Hole current from p-region to n-region in a p-n junction.

hill and slide back down into the p-type region. We shall not be concerned with holes which follow this particular course. On the other hand, it may, by chance, diffuse more deeply into the n-type region. In this event, it will on the average diffuse for lifetime τ, and subsequently it will be captured by a deathnium center in which it recombines with an electron.

The average depth to which holes diffuse in the n-type region depends upon the lifetime. The holes spread out in the region by diffusion. When the suitable differential equation describing this process is solved, it is found that the average depth to which they penetrate is given by the equation.

$$L = \sqrt{D\tau}$$

where L is known as the diffusion length, D is the diffusion constant for holes, and τ is the lifetime for holes in the n-region. Thus under equilibrium conditions a current of holes flows from the p-region into the n-region and penetrates on the average one diffusion length L before recombining with electrons.

Under equilibrium conditions a *principle of detailed balance* holds. This principle of statistical mechanics says that each process and its opposite occur equally frequently. Hence we must conclude that the flow of holes from the p-region into the n-region, followed by recombination, must be exactly balanced by a reverse process. The reverse process is thermal generation of holes through deathnium centers, followed by diffusion to barrier where they slide down into p-type region.

The application of voltage to the terminals of the device shown in Fig. 5 destroys the exact balance of the two currents just discussed. In considering the application of voltage we shall neglect any voltage drops at the contacts between the metal electrodes of Fig. 5 and the semiconductors. At the end of this section we will return briefly to the reasons why such voltage drops may be neglected. The effect of the application of voltages upon the currents is represented in Fig. 8. In part (*a*) of this figure we show the thermal equilibrium condition. The two currents previously discussed are represented by I_f and I_g, these currents standing, respectively, for the current of holes entering the n-region and recombining and the current generated in the n-region and diffusing to the barrier[21]. For the condition of thermal equilibrium these two currents are equal and opposite. In part (*b*) of the figure the situation for a large « reverse » bias is shown. For reverse bias, negative voltage is applied to the p-region and positive to the n-region so that the electro-

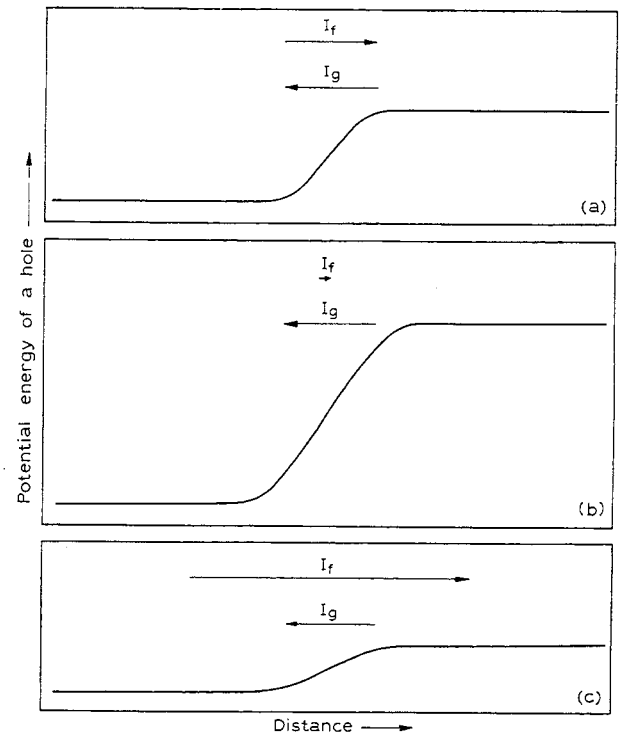


Fig. 8. Dependence of recombination and generation currents upon bias. (*a*) Thermal equilibrium; (*b*) reverse bias; (*c*) forward bias.

static potential difference between the two regions is increased. If the electrostatic potential is sufficiently high, corresponding to the situation shown in part (*b*), then practically no holes can climb the potential hill and I_f drops nearly to zero. This situation is represented by showing I_f as a vector of negligible length, whereas I_g has practically the same value as it has for the case of thermal equilibrium. In general, the diffusion length L is large compared to the width of the dipole or space-charge region. Hence the region where I_g arises is practically unaffected by reverse bias and I_g is thus independent of reverse bias. This independence of current upon bias is referred to as *saturation*.

When forward bias is applied, the situation shown in Fig. 8 (*c*) occurs and I_f increases. This increase is described in terms of the energy difference for a hole in the n-region and p-region. This energy difference is equal to the charge of the electron times the electrostatic potential differences be-

tween the two sides. We can apply a general theorem from statistical mechanics to a consideration of the number of holes which, by chance, acquire sufficient energy to climb the potential hill. This theorem states that each time the potential hill is increased by one thermal unit of energy, kT, then the number of holes capable of climbing the higher hill is reduced by a factor of 1/e. Since the potential barrier is already present under conditions of thermal equilibrium, it follows also that each lowering of the barrier by an amount kT will increase the current by a factor of e. The change in height of the barrier caused by the applied voltage V is $-qV$, where the polarity is so chosen that positive values correspond to plus potentials applied to the p-region and q is the absolute value of the charge of the electron. $V = 0$ is the equilibrium case, and for this case I_f is equal to I_g. Hence, in general, the recombination current is

$$I_f = I_g \exp qV/kT$$

This gives rise to a total current of holes from p-region to n-region, given by the difference

$$I_f - I_g = I_g [\exp (qV/kT) - 1]$$

This current is zero when $V = 0$, increases exponentially to large values for positive V, and decreases to a negative saturation value of I_g when V is negative and substantially larger than kT/q.

Similar reasoning can be applied to the electron current flowing across the junction. The applied potential which lowers the potential barrier for holes, evidently lowers it also for electrons; consequently, large electron currents flow under the same voltage conditions that produce large hole currents. In both cases these large currents correspond to flows of minority carriers into each region. In both cases the current is in the sense of transferring positive charge from the p-region to the n-region. In one case this is carried in the form of positive imperfections, namely holes, moving from p to n, and in the other case it is due to negative imperfections, electrons, moving in the opposite direction. For the case of reverse biases the potential rise is larger and the holes tend to be retained in the p-region and the electrons in the n-region. A saturation current due to generation in both regions flows. If the total saturation current is called I_s, then the total current for any applied voltage V is given by the formula

$$I = [\exp (qV/kT) - 1] I_s$$

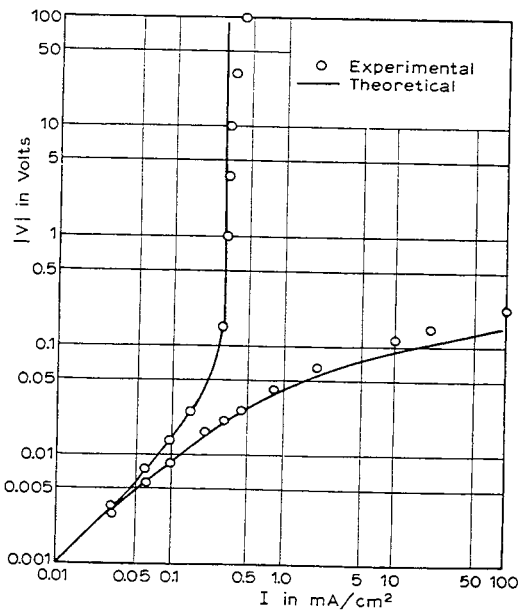


Fig. 9. Rectification characteristic for a p-n junction.

Evidently, I_s is the sum of the two generation currents. This equation is found to be well-satisfied for p-n junctions in germanium, and a comparison of the rectification curve[22] as measured with the theoretical formula is given in Fig. 9. It should be noted that the separation between the forward and reverse branches of the curves corresponds to a factor of e when the voltage is $kT/q = 25$ mV. This is exactly the factor predicted by the preceding equation. This agreement between theory and experiment is evidence that the imperfections which carry the current in a p-n junction have the charge of the electron. If they had twice this charge, a value of 12.5 mV should be obtained; if half the charge, the value should be 50 mV.

For large forward biases, the potential barrier between the n- and p-type is nearly eliminated. Under these conditions large concentrations of minority carriers flow across the barrier and density of the majority carriers may be substantially disturbed. Under these conditions it is no longer valid to consider that the minority carriers diffuse in a field-free region, as is the case when nearly equilibrium carrier densities exist. Although these large signal conditions are of general interest, we shall not consider them further here.

There are a number of ways in which the diffusion theory for rectification in the p-n junction can be tested experimentally. We shall consider a test

based upon a photoelectric effect. Photons of sufficient energy can generate hole-electron pairs when they are absorbed in germanium. This generation adds to the thermal generation produced in the deathnium centers. If the light is focused on the specimen in the form of a small spot, then it is possible to generate minority carriers in either region at a precise distance from the junction. The current flowing across the junction as a result of generation by the light should then decrease with increasing distance of the light from the junction. It can be shown that the probability that a minority carrier generated at a distance x from the junction diffuses to junction before recombining is simply

$$\exp(-x/L)$$

where L is the diffusion length of the minority carrier in the region where it is generated. This exponential dependence of response upon distance from the junction has been verified directly by Goucher and his colleagues[22]. They have found also that the value of L determined by studies of this sort is consistent with that necessary to explain value of I_s in the rectification formula.

For purposes of illustration we shall consider the value of the diffusion length for a typical example. A typical lifetime for a minority carrier is 10^{-4} sec and the diffusion constant for electrons in germanium is 93 cm^2/sec. These lead to a diffusion length of about 0.1 cm or 1 mm.

Since the light acts as a current generator, its effect may be readily included in an equivalent circuit for a p-n junction. This is illustrated in Fig. 10. Here I_l is the current of minority carriers generated by the light. The equivalent circuit shown corresponds to a case in which the light is focused at a distance x from the junction. If the light is distributed, the appropriate average of the exponential probability factor should be used. If the light falls on both the n-type and the p-type region, this average should take into account the fact that the diffusion length for holes in the n-region is probably different from that for electrons in the p-region. Fig. 10 emphasizes the importance of considering the light as acting as a current generator. If the equivalent circuit of Fig. 10 is operated in an open-circuit condition, then a photovoltage will be developed. This photovoltage will in general be nonlinear in the light intensity because of the rectification characteristics of the rectifier. In the equivalent circuit of Fig. 10, however, the current generator is linear in the light intensity. This independence of the current generator of biases applied to the p-n junction has been verified over a wide range of ex-

perimental conditions by Goucher[22] and his colleagues, and by Pietenpol[23].

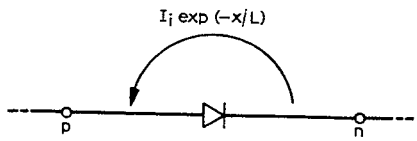


Fig. 10. Equivalent circuit for a p-n junction photocell.

The junction transistor[18,24,25]

The junction transistor is a compositional structure composed of two p-n junctions placed back to back. In its general behaviour as an amplifying device, a junction transistor shows great similarities to a vacuum-tube triode, or thermionic valve. Fig. 11 shows an n-p-n junction transistor in an amplifying circuit, the transistor being in the form of a sandwich with a layer of p-type germanium interposed between two layers of n-type germanium. Non-rectifying electrical contacts are made to the three layers. Under operating conditions the n-type region on the right, known as the collector, is biased positive so as to become attractive to electrons. As a result, a reverse bias appears between it and the middle region, known as the base. The current flowing across this reverse-biased junction can be controlled by an input signal applied between the base layer and the n-type region to the left, known as the emitter. As I shall describe below in more detail, the bias across the emitter junction controls the electron flow into the base region. In effect, the emitter junction acts like the region between the cathode and the grid in the vacuum tube. Electrons which enter the base region have a high probability of diffusing to the collector junction, and thus the flow of electrons from emitter to collector can be varied by varying the potential across the emitter base junction. The action is very similar to that controlling the flow of electrons from the cathode to the anode in the thermionic triode.

Junction transistors can be fabricated in a variety of ways. The compositional structure can be produced in a crystal-growing machine by techniques like those used for making simple p-n junctions. As the crystal is grown from a melt containing antimony, a pellet containing indium is dropped into the melt, and a second pellet containing antimony is dropped in a few seconds later. The portion of the crystal which grows between the dropping of the two pellets is rich in indium and is consequently p-type. The second pellet

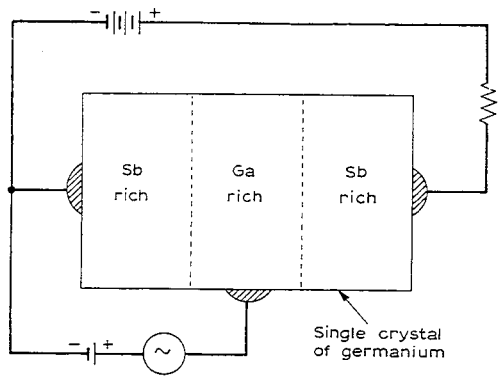


Fig. 11. The structure of a junction transistor and the bias supply for its operation in an amplifying circuit.

over-compensates the effect of the added indium and the subsequent material is again n-type. From such a single crystal small rods may be cut and contacts made. There are a number of technical processes involved in proceeding from the point of growing the crystal to the production of a packaged, stable transistor. I shall not endeavor to discuss these in the lecture.

An alternative technique for producing the compositional structure[25a] starts with a thin plate of germanium which subsequently plays the role of the p-type region. A pellet of metal containing a donor is placed on this plate. The plate and pellet are then raised to such a temperature that the metal melts and dissolves a small amount of germanium. When the metal and germanium are subsequently cooled, the germanium precipitates from the metal and grows back onto the crystal structure of the base material. This regrown germanium carries with it some of the donors contained in the molten metal and thus grows an n-type region. In the fabrication of a transistor, pellets are placed on both sides of a thin plate and allowances are made for the degree to which they dissolve germanium on the two sides. This process has been used in the production of a large fraction, if not the majority, of transistors made to date.

From the point of view of an electron the situation in an operating transistor is as represented in Fig. 12. This diagram shows the variation in potential energy for an electron along a line going from emitter to collector in a transistor biased like that shown in Fig. 11. The reverse bias at the collector junction produces a large drop in potential on the right-hand side. The varying bias across the emitter junction changes the height of the hill and thus

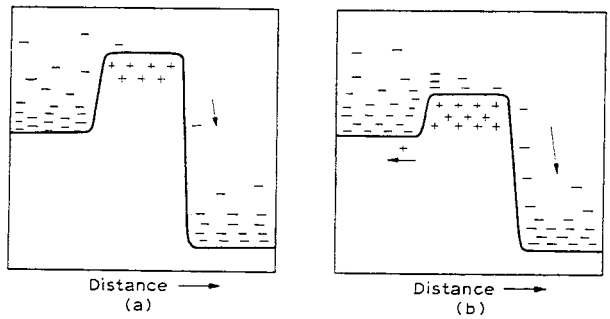


Fig. 12. The potential energy of an electron in an n-p-n junction transistor for two values of bias across the emitter junction. In (*b*) the forward bias is greater than in (*a*).

varies the diffusion current of electrons into the base. The base layer is very thin and contains very little deathnium. Consequently, the probability is much higher for an electron to diffuse through the base layer and arrive at the collector junction than for it to combine with a hole through a deathnium center in the base layer. In a well-built junction transistor, in fact, the electron flow through the base region proceeds so efficiently that the electron current flowing through the base layer to the collector may exceed the current combining in the base layer by a factor of 100 or more. This means that the input currents flowing to the base layer through the base lead may control the output currents 100 times larger flowing to the collector region. Although this situation is not nearly as ideal as in a vacuum triode, in which the grid current is still smaller compared with the anode current, it permits the junction transistor to operate as a highly efficient amplifier.

So far as voltage requirements are concerned, the junction transistor is far superior to any vacuum tube, because a junction transistor can be brought fully into its operational range with voltages as small as 50–100 mV. Such voltages make the potential hill at the collector junction shown on Fig. 12 several times the thermal voltage ($kT/q = 25$ mV). As a consequence, any electron reaching the collector junction is sure to slide down the hill and has a negligible chance of returning.

The junction transistor has almost ideal pentode characteristics in terms of collector saturation. This also follows from the same considerations which make it operate well at low collector voltages. In fact, in some junction transistors there is a negligible variation in current over a voltage range of a hundredfold or more, say from 100 mV to 10 volts on the collector.

The junction transistor described in Figs. 11 and 12 is only one of a

large family of transistors. The p-n-i-p and n-p-i-n transistors are variations of this form especially designed for low capacitance between the collector and the base[25b]. The so-called junction tetrode is a special form of junction transistor in which the current flow is controlled so as to occur only over a small region of the base[25c]. Junction transistors of these types have been used in oscillating circuits at frequencies as high as 1000 Mc/s.

In closing this section, I should like to point out that the junction transistor operates more flexibly in terms of power than do vacuum tubes. This is of great importance, since it means that junction transistors may be used efficiently in some cases where the power level to be amplified is very much smaller than the heater current required in the most efficient of vacuum tubes. The possibility of operating at very low power is due to the low voltages necessary to operate a junction transistor, and because of its small size and the high degree of purity possible, the current due to generation by deathnium centers can be made smaller than a microampere. It is thus possible to make a junction transistor which can be put into an amplifying condition with a power input substantially less than a microwatt. The same transistor, however, can operate at voltages as high as 10 volts and currents of the order of 10 mA. Thus, it can cover a power range effectively of 100,000. This fact indicates that there will probably be a much smaller diversity of transistor types than there are of vacuum tube types.

Conclusion

In a lecture of reasonable scope, it is no longer possible to do justice to the many branches of semiconductor physics that have developed since the discovery of the transistor effect. In the previous sections I have tried to illustrate how the stimulation of aiming for practical goals led to an experimental program which brought out many new and important aspects of electronic phenomena in semiconductors.

Not all of the important work has, of course, been motivated by highly practical ends which could be so immediately achieved once the fundamental science was understood. Much of the work on surface states[8,26] should now be regarded as being in the pure-science class inasmuch as it is not possible to see at all directly how the scientific results will find utility. However, it seems highly probable that once the phenomena of surface states are thoroughly understood from a scientific point of view, many useful sug-

gestions will arise as to how this knowledge may be employed to make better devices.

Also in the class of purely scientific work are the very beautiful determinations of the energy band structure in silicon and germanium that have resulted from *cyclotron resonance* in silicon and germanium[27,28,29]. This work has led to a more detailed knowledge of energy band surfaces than has ever existed before, except possibly for some of the most simple and almost trivial cases.

One of the large new areas that has begun to grow as a result of the large amount of transistor technology may be called « The Theory of Solid Solutions in Silicon and Germanium ». Recently, interactions between donors and acceptors of various types have been given active consideration[30]. It is evident that interesting phenomena are associated with the incorporation of any foreign element from the Periodic Table in silicon and germanium. Important distinctions are being drawn between the same atoms in a substitutional position and an interstitial position. It is also being found that some atoms may exist in a semiconductor in not merely two states of charge as in the case of a simple donor and acceptor but in as many as four different conditions of charge[31]. It seems highly probable that this area of work will become a field of active interest in physical chemistry in the years to come.

Among the phenomena motivated in considerable measure by practical aims we should mention the phenomena of avalanche breakdown which is useful in voltage limiting diodes and protection devices. A strong motivation for following up early leads which suggested that such effects might occur, came from the conversation I once had with Dr. O. E. Buckley, whom I have mentioned before in this lecture. Dr. Buckley pointed out the need for a device for protecting telephones from damage due to voltages induced by lightning near telephone lines. It was in large measure my knowledge of this need that gave emphasis to this work. The original interpretation[32] which appeared to be consistent with the dielectric breakdown mechanism once proposed by Zener[33] (direct rupture of electron-pair bonds by strong electric fields with the resultant production of hole-electron pairs) seems now to have been in error. The investigation of the same phenomena by McKay[34] and his collaborators has led to a new branch of semiconductor electronics dealing with impact ionization of holes in electrons moving in intense electric fields in semiconductors. Both the effect and its theory are playing an active role in a certain class of switching transistors.

In closing this lecture, I would like to refer to a paragraph written in my

book in 1950[35]. I am pleased to see that the predictions of the paragraph appear to have been borne out to a considerable extent and I feel that it is now as appicable as it was then:

«It may be appropriate to speculate at this point about the future of transistor electronics. Those who have worked intensively in the field share the author's feeling of great optimism regarding the ultimate potentialities. It appears to most of the workers that an area has been opened up comparable to the entire area of vacuum and gas-discharge electronics. Already several transistor structures have been developed and many others have been explored to the extent of demonstrating their ultimate practicality, and still other ideas have been produced which have yet to be subjected to adequate experimental tests. It seems likely that many inventions unforeseen at present will be made based on the principles of carrier injection, the field effect, the Suhl effect, and the properties of rectifying junctions. It is quite probable that other new physical principles will also be utilized to practical ends as the art develops.»

It is my hope to contribute to the fulfilment of the predictions of this paragraph through my new organization in California.

1. J. Bardeen, *Nobel Lecture*, this volume, p. 318.
1a. J. H. Scaff, H. C. Theurer, and E. E. Schumacher, *Trans. AIME*, 185 (1949) 383.
2. W. Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand Co., Inc., New York, 1950, Ch. 12.
3. K. Lark-Horovitz, *Semi-Conducting Materials*, Butterworth Scientific Publications, London, 1951, pp. 47–48.
4. J. A. Burton, G. W. Hull, F. J. Morin, and J. C. Severiens, *J. Phys. Chem.*, 57 (1953) 853–859.
5. The theory of this process has been developed by W. Shockley and W. T. Read, Jr., in *Phys. Rev.*, 87 (1952) 835–842. Experimental findings in agreement with the theory have been obtained by Burton, Hull, Morin, and Severiens (Ref. 4) and by R. N. Hall, *Phys. Rev.*, 83 (1951) 228; *Ibid.*, 87 (1952) 387.
6. J. Bardeen, *Phys. Rev.*, 71 (1947) 717–727.
7. W. Shockley and G. L. Pearson, *Phys. Rev.*, 74 (1948) 232–233.
8. W. H. Brattain, *Nobel Lecture*, this volume, p. 377.
9. J. N. Shive, *Phys. Rev.*, 75 (1949) 689–690.
10. J. R. Haynes and W. Shockley, *Phys. Rev.*, 75 (1949) 691.
11. J. R. Haynes and W. Shockley, *Phys. Rev.*, 81 (1951) 835–843.
J. R. Haynes and W. Westphal, *Phys. Rev.*, 85 (1952) 680.
12. Transistor Teacher's Summer School, *Phys. Rev.*, 88 (1952) 1368–1369.

13. H. Krömer, *Z. Physik*, 134 (1953) 435–450.
14. W. Shockley, *Bell System Tech. J.*, 33 (1954) 799–826.
15. E. J. Ryder and W. Shockley, *Phys. Rev.*, 81 (1951) 139–140.
16. E. J. Ryder, *Phys. Rev.*, 90 (1953) 767–769.
17. W. Shockley, *Bell System Tech. J.*, 30 (1951) 990–1034.
18. W. Shockley, *Bell System Tech. J.*, 28 (1949) 435.
19. G. K. Teal and J. B. Little, *Phys. Rev.*, 78 (1950) 647.
20. W. G. Pfann, *Trans. AIME*, 194 (1952) 747; W. G. Pfann and K. M. Olsen, *Phys. Rev.*, 89 (1953) 322; *Bell Lab. Record*, 33 (1955) 201.
21. The subscripts *f* and *g* may be thought of as *forward* and *generation*. This mixed choice avoids subscript *r*, which might be either *reverse* or *recombination*. In this decision, *forward* is equivalent to *recombination* and *generation* to *reverse*.
22. F. S. Goucher *et al.*, *Phys. Rev.*, 81 (1951) 637.
23. W. J. Pietenpol, *Phys. Rev.*, 82 (1951) 120.
24. W. Shockley, M. Sparks, and G. K. Teal, *Phys. Rev.*, 83 (1951) 151.
25. W. J. Pietenpol and R. L. Wallace, *Proc. I.R.E.*, 39 (1951) 753.
25a. R. N. Hall and W. C. Dunlap, *Phys. Rev.*, 80 (1950) 467.
25b. J. M. Early, *Bell System Tech. J.*, 33 (1954) 517.
25c. R. L. Wallace, L. G. Schimpf, and E. Dickten, *Proc. I.R.E.*, 40 (1952) 1395.
26. For a recent review and references, *see* R. H. Kingston (Ed.), *Semiconductor Surface Physics*, University of Pennsylvania Press, Philadelphia, 1957.
27. W. Shockley, *Phys. Rev.*, 90 (1953) 491.
28. G. Dresselhaus, A. F. Kip, and C. Kittel, *Phys. Rev.*, 92 (1953) 827. R. N. Dexter, B. Lax, A. F. Kip, and G. Dresselhaus, *Ibid.*, 96 (1954) 222.
29. R. N. Dexter, H. J. Zeiger, and B. Lax, *Phys. Rev.*, 104 (1956) 637–644 (references).
30. H. Reiss, C. S. Fuller, and F. J. Morin, *Bell System Tech. J.*, 35 (1956) 535–636.
31. H. H. Woodbury and W. W. Tyler, *Phys. Rev.*, 105 (1957) 84.
32. K. B. McAfee, E. J. Ryder, W. Shockley, and M. Sparks, *Phys. Rev.*, 83 (1951) 650.
33. C. Zener, *Proc. Roy. Soc. London*, A 145 (1934) 523.
34. K. G. McKay, *Phys. Rev.*, 94 (1954) 877.
35. W. Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand Co., Inc., New York, 1950, p. 349.

Biography

William Shockley was born in London, England, on 13th February, 1910, the son of William Hillman Shockley, a mining engineer born in Massachusetts and his wife, Mary (*née* Bradford) who had also been engaged in mining, being a deputy mineral surveyor in Nevada.

The family returned to the United States in 1913 and William Jr. was educated in California, taking his B.Sc. degree at the California Institute of Technology in 1932. He studied at Massachusetts Institute of Technology under Professor J. C. Slater and obtained his Ph.D. in 1936, submitting a thesis on the energy band structure of sodium chloride. The same year he joined Bell Telephone Laboratories, working in the group headed by Dr. C. J. Davisson and remained there (with brief absences for war service, etc.) until 1955. He resigned his post of Director of the Transistor Physics Department to become Director of the Shockley Semi-conductor Laboratory of Beckman Instruments, Inc., at Mountain View, California, for research development and production of new transistor and other semiconductor devices. In 1963 he was named first Alexander M. Poniatoff Professor of Engineering Science at Stanford University, where he will act as professor-at-large in engineering and applied sciences.

During World War II he was Research Director of the Anti-submarine Warfare Operations Research Group and he afterwards served as Expert Consultant in the office of the Secretary for War.

He held two visiting lectureships: in 1946 at Princeton University, and in 1954 at the California Institute of Technology. For one year (1954–1955) he was Deputy Director and Research Director of the Weapons System Evaluation Group in the Defence Department.

Shockley's research has been centred on energy bands in solids; order and disorder in alloys; theory of vacuum tubes; self-diffusion of copper; theories of dislocations and grain boundaries; experiment and theory on ferromagnetic domains; experiments on photoelectrons in silver chloride; various topics in transistor physics and operations research on the statistics of salary and individual productivity in research laboratories.

His work has been rewarded with many honours. He received the Medal for Merit in 1946, for his work with the War Department; the Morris Leibmann Memorial Prize of the Institute of Radio Engineers in 1952; the following year, the Oliver E. Buckley Solid State Physics Prize of the American Physical Society, and a year later the Cyrus B. Comstock Award of the National Academy of Sciences. The crowning honour – the Nobel Prize for Physics – was bestowed on him in 1956, jointly with his two former colleagues at the Bell Telephone Laboratories, John Bardeen and Walter H. Brattain.

In 1963 he was selected as recipient of the Holley Medal of the American Society of Mechanical Engineers.

Dr. Shockley has been a member of the Scientific Advisory Panel of the U.S. Army since 1951 and he has served on the Air Force Scientific Advisory Board since 1958. In 1962 he was appointed to the President's Scientific Advisory Committee. He has received honorary science doctorates from the University of Pennsylvania, Rutgers University and Gustavus Adolphus Colleges (Minn.).

In addition to numerous articles in scientific and technical journals, Shockley has written *Electrons and Holes in Semiconductors* (1950) and has edited *Imperfections of Nearly Perfect Crystals* (1952). He has taken out more than 50 U.S. patents for his inventions.

Dr. Shockley has been married twice, and has three children by his first marriage to Jean (*née* Bailey). This union ended in divorce; his second wife is Emmy Lanning.

WALTER H. BRATTAIN

Surface properties of semiconductors

Nobel Lecture, December 11, 1956

First let me say that while I am very proud to be one of the recipients of the Nobel Award in Physics, I am nevertheless well aware that I am only a representative of many others, without whose work and effort I would not be here today. May I mention first my teachers and in particular the late Prof. Benjamin H. Brown at Whitman College, who originally inspired me to take up physics and who also taught my mother and father. Then there are all the many workers in the field of semiconductors on whose shoulders we have stood. Finally and more closely connected with our work are all of our colleagues at the Bell Telephone Laboratories whose efforts and cooperation made our work possible. In particular here I would like to mention R. S. Ohl, J. H. Scaff and H. C. Theuerer, whose pioneering work on silicon made a new class of semiconductors available to physicists. It was Scaff and Theuerer who, as metallurgists, supplied the silicon and germanium with which we worked, and Ohl who first discovered a p-n junction in a melt of silicon prepared for him by the other two. One needs indeed to be very humble about accepting such an award when he thinks how fortunate he was to be in the right environment at the right time, to take advantage of all that had been done before. Let me here express my gratitude to all of these people.

I would like to start by emphasizing the importance of surfaces. It is at a surface where many of our most interesting and useful phenomena occur. We live for example on the surface of a planet. It is at a surface where the catalysis of chemical reactions occur. It is essentially at a surface of a plant that sunlight is converted to a sugar. In electronics, most if not all active circuit elements involve non-equilibrium phenomena occurring at surfaces. Much of biology is concerned with reactions at a surface.

If surfaces are so important, what do we know about them? What is a surface? What properties does a surface have that a physicist can measure? Of the various states of matter – solid, liquid, and gaseous – physicists have long understood and explained the behavior of gases and are now, thanks to quantum mechanics, making considerable progress with simple homoge-

neous solids. Liquids are still quite obscure and surfaces still more so. It was Irving Langmuir, a previous Nobel Laureate, who gave us the first model of a surface, namely clean tungsten in high vacuum, and who further explored some of the simple adsorption phenomena that take place at such a surface*. The main physical measurement by which he could characterize his surface was the electronic work function or potential jump at the tungsten surface and the changes in this potential due to adsorption of various foreign atoms or molecules. He could also measure heats of evaporation of these components. All this points up the essential properties of a surface, namely: (*i*) it is a boundary across which the concentration of one or more components changes greatly, and (*ii*) there is a potential jump or energy change associated with the surface. In the case of a metal all these phenomena take place in a few ångströms, making such a surface difficult to analyze. The behavior of a tungsten surface has for some time been a model for the understanding of all surface phenomena. When one comes to some of the dirtier and often more interesting surfaces these have been too remote from clean tungsten for it to serve well in understanding their behavior.

The main point of this story is that now as a result of the progress that has been made in the understanding of semiconductors we have another model of a surface, namely a p-n junction in a single crystal of a semiconductor such as silicon or germanium, and that this model is in a sense at the opposite extreme from clean tungsten. It can also be said that most surfaces lie in between these two models and either model may be used for understanding the multitude of surface phenomena, whichever is best.

In particular one can understand the complicated and dirty germanium gas surface at least qualitatively in terms of the p-n junction model. Let us first consider our p-n junction model. It occurs in one of the simplest, purest and now best understood solids. For example a single crystal of germanium, its atoms held together by well-understood covalent bonds, serves as a medium in which electrons and holes can exist in equilibrium. The existence of these components depends on the thermal dissociation of valence bonds namely

$$(\text{valence bond}) \rightleftharpoons n + p$$

* The present author got his start working under J. A. Becker who was one of the early contributors and is still actively working in the field of metal surfaces.

The law governing this reaction is the mass action law

$$np = K \tag{1}$$

where n and p represent the concentration of electrons and holes in equilibrium at a given temperature. The constant K depends primarily on the absolute temperature (T) and energy (E_g) necessary to dissociate a valence bond creating a hole electron pair:

$$K \sim \exp - (E_g/kT) \tag{2}$$

where k is Boltzmann's constant. The proportionality constant in this last equation can be approximately derived from fundamental physical constants and the density of the crystal. The parameter characterizing the semiconductor is the energy E_g necessary to dissociate a hole electron pair. It is approximately three quarters of an electron volt for germanium, one electron volt for silicon and ranges in other semiconductors from a few tenths to several electron volts.

The main point here is that while the product of the equilibrium concentrations is always a constant at a given temperature, the magnitude of either concentration can be varied over extreme ranges by solution of impurities in the semiconductor lattice. Substitional solution of a fifth-column element in the lattice of silicon or germanium, as discussed by Shockley in the previous lecture, makes n very large and p correspondingly small, the product remaining constant. Likewise the solution in the lattice of a third-column element makes p large and n small.

Consider now a p-n junction, a crystal of germanium, one side of which is n-type ($n_1 > p_1$) and the other side p-type ($p_2 > n_2$). At the contact between the two regions we have the situation that n_1 must be greater than n_2 and likewise p_2 greater than p_1 since $n_1 p_1 = n_2 p_2 = K$. Because of this concentration difference, electrons will tend to diffuse from 1 to 2 and holes from 2 to 1. This will charge region 1 positive and region 2 negative until an electric potential difference is built up just sufficient to make the field flow balance the diffusion flow. The crystal will then be in equilibrium. This potential is given by

$$V = (kT/e) \ln (n_2/n_1) = (kT/e) \ln (p_1/p_2) \tag{3}$$

Associated with this potential is a space-charge double layer at the boundary between the two regions. The charge density ϱ is related to the potential by Poisson's equation

$$\frac{d^2V}{dx^2} = -4\pi\varrho/\varepsilon \tag{4}$$

where ε is dielectric constant of the medium. The extent of this space-charge double layer is large compared to that in a metal. In germanium it has a characteristic length of about 1×10^{-4} cm.

This is our model of a surface, a phase boundary across which the concentration of the components differ greatly, the associated potential jump at the surface and the necessary charge double layer. Its properties can best be studied by changing the concentration of the components on each side of the surface from their equilibrium values. When one does this, one finds that the potential difference across the surface changes in a manner consistent with Eq. (3) and this change in potential can be measured. If p_1' and n_2' are the non-equilibrium concentrations of holes and electrons on sides 1 and 2 then the change in potential is given by

$$\Delta V = (kT/e) \ln (p_1'/p_1) = (kT/e) \ln (n_2'/n_2) \tag{5}$$

Note that it is the change in concentration of the minority carrier on each side of the boundary that is controlling factor. Likewise, by changing from the equilibrium to the non-equilibrium state, one can measure the flow of charge into or out of the sample or the corresponding change in charge at the double layer. By a consideration of rate processes going on at this interface, one can also write down a current balance equation for this surface.

$$I = I_s \left[\frac{p_1'}{p_1} - 1\right] \tag{6}$$

where I_s is the so-called saturation current and is determined by rates of recombination and thermal generation of electrons and holes on both sides of the surface. The non-equilibrium condition may be obtained by flow of current across the surface, in which case I is this current, or by generation of electron hole pairs in the neighborhood of the junction, for example by light, in which case I is the rate of generation of pairs by light multiplied by

the electron charge (e). If both means are used then I is the sum of both currents. Note that one can eliminate p_2'/p_1 from Eqs. (5) and (6), getting the well-known rectifier equation.

We see that the physical properties of this surface which can be measured are the changes in potential, changes in concentration occurring at the surface, and rates of return to equilibrium. These are in general the things one can measure for any surface.

Let us now consider a germanium surface in a gaseous ambient. By the Kelvin method one can measure a contact potential difference between it and a reference electrode. Any change in the potential jump at the surface will give a corresponding change in the contact potential. If the equilibrium at the surface is upset by shining light on the surface the resulting potential change can be measured. This is the surface photo-effect. If the germanium is the proper shape, thin and long, one can also measure the photo conductivity, and from this deduce the change in minority carrier concentration for a given illumination. On a large sample of germanium one can measure the lifetime of electrons and holes in the body of the germanium and then, from the transient behavior of the thin sample, the surface rate of recombination of holes and electrons. As a result of all these studies, one finds that there is in general a space-charge double layer at the germanium surface. It is as if the germanium surface was of a given type, p or n, depending on the nature of the ambient gas and that the germanium interior must be in equilibrium with its surface. All the qualitative phenomena that one would find at a p-n junction, if one maintained say side 1 at a given fixed n-type and varied side 2 from p^+ to n^+, are found at a germanium gas interface. For example in the case of an n-(p^+) junction the photo e.m.f. is such to make the n-side more negative. As one changes the p^+ side from p^+ to just the same degree of n-ness as side 1, the photo e.m.f. goes to zero and when side 2 is made n^+ the photo e.m.f. changes sign. This e.m.f. is largest in magnitude for n-(p^+) case. If side 1 had been p-type, similar changes would occur when side 2 was varied from p^+ to n^+ but now the photo e.m.f. is largest for the p-(n^+) case. This is also quite an accurate description of the germanium-gas surface. Moreover the magnitude and sign of the photo effects at the extremes can be predicted if one knows the density of holes and electrons, the body lifetime and the surface recombination for the sample in question. Furthermore the experimental results agree with theory.

From the above results one would predict that, since the surface of a germanium crystal in a gaseous ambient may be of a different type from the

interior, the electrical conductivity of the surface should be different from the interior, and, as the surface changes in type, this conductivity should change. In a thin-enough sample these changes can be measured. Along with the changes in surface type, there must also be a change in the charge double layer. One can of course induce such a change in the charge double layer by applying an electrical field perpendicular to the surface, and one can measure the change in conductivity induced by this field. This is the field effect which was predicted by Shockley but which at first one was unable to detect.

As a result of measurements of this change of conductivity with field and of the other changes at the surface with change in gaseous ambient, it is found that there are slow and fast changes at the surface. For example, if one illuminates the surface of germanium suddenly, the contact potential changes quickly to a new value, then slowly decays back towards the original value. Likewise with the surface conductivity when one suddenly applies the field. There are of course changes in the charge double layer at the surface corresponding to the change in surface photo-effect and surface conductivity. From these results one gets the concept of slow and fast states at the surface wherein this charge resides. The fast states are so to speak in good connection with the germanium body and can get into equilibrium or steady state with the body very quickly, in a microsecond or less. On the other hand the slow states are poorly connected with the interior. It takes seconds or minutes for them to adjust to changes induced by light or field. These may not even be states at all but changes in adsorption equilibrium or steady state with the surrounding gas.

The next step is that, from our knowledge of the properties of germanium, one can calculate, from the changes in conductivity, the potential difference across the space layer from the surface into the interior. One can also calculate the change in charge in the space-charge layer. As one changes the germanium surface from p^+ to n^+, the conductivity should go through a minimum. If in an experiment one can find this minimum, one knows the potential difference across the space-charge layer and, from the change in conductivity from this minimum, one knows the change in this potential. Further, from the field-effect experiment, one can determine the total change in charge at the surface, and from the change in conductivity, how much of this change in charge occurred in the space-charge layer. The charge across the surface must of course add up to zero when there is no field at the surface. If Σ_b, Σ_f and Σ_s are the charges in the space-charge barrier layer, the fast states and the slow states respectively, then

$$\Sigma_b + \Sigma_f + \Sigma_s = 0$$

In the case of an applied field, the sum of these charges must be equal to the field-induced charge. By calculating Σ_b and measuring changes in total charge under slow and fast conditions one can obtain values for both Σ_f and Σ_s. One can also determine the potential change across the space-charge layer and the total change in potential across the surface. For this type of surface one has the chance of taking the surface apart and measuring the individual parts in some detail. What remains to be done is to determine at the same time the chemical changes that occur at this surface and to understand the relation between the physical and chemical changes. When this is accomplished – and I think it will be soon – the germanium surface will be the most thoroughly understood of all surfaces.

As an example of what already has been accomplished if one measures on the same surface at the same time certain quantities as a function of gas ambient, one can deduce the distribution in energy and capture cross-sections of the fast states or traps for electrons and holes. These quantities are the photoconductivity, the decay lifetime, the change of conductivity with field and the surface photo-effect. From these measurements and a knowledge of the body properties of the germanium sample one can then deduce (*i*) the change of charge in the fast states with respect to the surface space-charge layer potential difference at zero added carrier concentration, (*ii*) the change of charge in the fast states with added carrier concentration at constant surface potential, and (*iii*) the surface recombination, all as a function of ambient gas or surface potential. These quantities all depend on the fast trap distribution in energy and on the surface potential. The second depends also on the ratio of the capture cross section for holes and electrons and the third depends as well on the product of the capture cross sections. The relations are integral equations and not easy to solve in general. One question is whether the trap distribution is discrete or continuous. An examination of all the pertinent data indicates that, at least for energies near the middle of the region between the valence and conduction bands, the distribution of traps in energy is continuous and that it is fixed and independent of gaseous ambient or germanium type, for a given surface treatment. One can calculate approximately the ratio of the cross sections obtaining $\sigma_p/\sigma_n \cong 150$ and the product $\sigma_p \times \sigma_n \cong 2.4 \times 10^{-31}$ cm^4 giving $\sigma_p \cong 6 \times 10^{-15}$ cm^2, $\sigma_n = 4 \times 10^{-17}$ cm^2. Not only are these values reasonable but they suggest that the fast trap is acceptor-like, that is, negatively charged when occupied by an

electron and neutral when empty. The attraction between the hole and the negatively charged trap thus explains the larger cross section for hole capture.

In conclusion, one can say that the physical nature of the germanium surface in a gas depends primarily on the surface treatment and the nature of the gas, and not on the type, p or n, of the body material. The reaction of the surface with the gaseous ambient, in particular the Bardeen-Brattain cycle of oxygen and water vapor, is an example of a low-grade catalytic reaction. The understanding of such a simple surface will ultimately contribute to understanding of other surface phenomena, especially catalysis. It was the original attempts to understand surface phenomena of this nature that led to the discovery of the transistor effect. Since then many people have contributed to the present understanding. Some of these are the work of Bardeen and his group at the University of Illinois, Henisch and his group at the University of Reading, Kingston and his co-workers at Lincoln Laboratory, Aigrain and Dugas at the École Normale, Many and his group at the Hebrew University, Israel, and that of my colleague – Garrett. Equally important is the work of Walter Brown on the field effect and, while not specifically dealt with here, his earlier work on channel conduction along with the similar work of Statz and his group at Raytheon.

Biography

Walter H. Brattain was born in Amoy, China, on February 10, 1902, the son of Ross R. Brattain and Ottilie Houser. He spent his childhood and youth in the State of Washington and received a B.S. degree from Whitman College in 1924. He was awarded the M.A. degree by the University of Oregon in 1926 and the Ph.D. degree by the University of Minnesota in 1929.

Dr. Brattain has been a member of the Bell Laboratories technical staff since 1929. The chief field of his research has been the surface properties of solids. His early work was concerned with thermionic emission and adsorbed layers on tungsten. He continued on into the field of rectification and photo-effects at semiconductor surfaces, beginning with a study of rectification at the surface of cuprous oxide. This work was followed by similar studies of silicon. Since World War II he has continued in the same line of research with both silicon and germanium.

Dr. Brattain's chief contributions to solid state physics have been the discovery of the photo-effect at the free surface of a semiconductor; the invention of the point-contact transistor jointly with Dr. John Bardeen, and work leading to a better understanding of the surface properties of semiconductors, undertaken first with Dr. Bardeen, later with Dr. C. G. B. Garrett, and currently with Dr. P. J. Boddy.

Dr. Brattain received the honorary Doctor of Science degree from Portland University in 1952, from Whitman College and Union College in 1955, and from the University of Minnesota in 1957. In 1952 he was awarded the Stuart Ballantine Medal of the Franklin Institute, and in 1955 the John Scott Medal. The degree at Union College and the two medals were received jointly with Dr. John Bardeen, in recognition of their work on the transistor.

Dr. Brattain is a member of the National Academy of Sciences and of the Franklin Institute; a Fellow of the American Physical Society, the American Academy of Arts and Sciences, and the American Association for the Advancement of Science. He is also a member of the commission on semi-

conductors of the International Union of Pure and Applied Physics, and of the Naval Research Advisory Committee.

In 1935 he married the late Dr. Keren (Gilmore) Brattain; they had one son, William Gilmore Brattain. In 1958 he married Mrs. Emma Jane (Kirsch) Miller. Dr. Brattain lives in Summit, New Jersey, near the Murray Hill (N.J.) laboratory of Bell Telephone Laboratories.

Physics 1957

CHEN NING YANG

TSUNG DAO LEE

«for their penetrating investigation of the so-called parity laws which has led to important discoveries regarding the elementary particles»

Physics 1957

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

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

The Nobel Prize in Physics to Professor Tsung Dao Lee and Professor Chen Ning Yang for this year is concerned with some of the fundamental physical principles, the so-called parity laws – in the first place the symmetry of Nature with respect to right and left – in their application to elementary particles and their reactions.

When during this century the old dream about atoms came true, it soon became clear not only that reality is by far richer than was the dream but also considerably different from it. The atoms that were found and which one learnt to count and to measure did by no means correspond to the ideal of indivisibility and unchangeability of the old atomists. But instead, there appeared a hitherto unknown, strangely unvisualizable feature of indivisibility in all atomic processes and therefore also behind all usual physical events, without which everything in the world would be in flux. The new edifice of laws, which was the consequence of these discoveries – it is called quantum theory – contains the laws of earlier physics as a correct but greatly simplified limiting case. They have a similar relationship to the laws of atomic physics as an airphoto to a near-picture of the same landscape.

The lesson learnt from quantum theory made that the literally unchangeable atoms of the old philosophers were hardly seriously sought in those particles – electrons, protons, and neutrons – from which atoms are made. As the name elementary particles under which they are summarized would seem to indicate, there was nevertheless a certain inclination towards this direction. But already the ordinary elementary particles are by no means unchangeable, which is still more strongly the case with the lot of new, similar particles discovered during later years, the transformations of which now stand at the centre of interest of atomic physicists. In our attempts to find a theory which comprehends all the new facts concerning old and new elementary particles, certain wittingly unreal, symbolic particles appear in our equations, which with a little good will could be regarded as the eternally immutable atoms of philosophers. Of the real elementary particles we could

then, following Lao-tse, the old Chinese thinker, say: « The elementary particles, which could be defined, are not the eternal elementary particles. » Lao-tse did not talk of elementary particles, of course, but of Tao, the deepest principle of life. And physics is certainly considerably simpler than human life. And there we have powerful auxiliaries in the art of experimentation and mathematics.

As to mathematics and elementary particles it has, in the first place, led to two theories, each of which has been developed by a Nobel Prize winner in Physics, Dirac and Fermi. Hereby the former theory is the outermost wing of the edifice of quantum mechanics, while the latter may be regarded as the first, still unfinished room in the new edifice of elementary particle laws. But they are both concerned with electrons and thus border on one another.

But what has the question of right and left to do with elementary particle physics? Well, in the first place only in a negative way, in that it was assumed almost tacitly, that elementary particle reactions are symmetric with respect to right and left. This assumption was to play an important part in the elaboration of Fermi's theory. That this assumption was made was very natural, not least in view of the mentioned theory of Dirac, according to which it looked as if the electrons, the best known elementary particles, possessed no feature which would permit a distinction between right and left. In fact, most of us were inclined to regard the symmetry of elementary particles with respect to right and left as a necessary consequence of the general principle of right-left symmetry of Nature. Thanks to Lee and Yang and the experimental discoveries inspired by them we now know that this was a mistake.

The starting-point of Lee and Yang in their revision of the whole question of right-left symmetry in elementary particle reactions were certain strange observations concerning a kind of new particles called K mesons, which looked as if they were in contrast with the assumption mentioned. Even if these observations puzzled greatly many physicists, it was only Lee and Yang who seriously took the consequences of them, in that they asked themselves what kind of experimental support there was for the assumption that all elementary particle processes are symmetric with respect to right and left. The result of their investigation was unexpected, namely that the validity of the symmetry assumption even in the best known processes had no experimental support whatsoever, the reason being that all experiments had been so arranged as to give the same result whether the assumption was valid or not. As if one had thought that Olav Tryggveson had his heart in the middle of

the body because he was equally skilled with the left as with the right hand.

Lee and Yang did not confine themselves to this negative statement but devised a number of experiments which would make it possible to test the right-left symmetry in different elementary particle transformations, and proposed them to their experimental colleagues. The first of these experiments was carried out by the Chinese physicist, Mrs. C. S. Wu and her collaborators. Very schematically it consisted in the following. Atomic nuclei of a radioactive isotope of the metal cobalt were exposed at very low temperature to a magnetic field – they are themselves small magnets – whereby they became directed just like compass needles. The distribution as to direction of the electrons due to radioactivity was then investigated.

Let us assume that the magnetic field is created by means of a coil placed like a spool of thread on a table, and that the electric current is flowing counterclockwise in the wire. Then the north poles of the cobalt nuclei will be directed upwards. The experiment, now, gave the result that the electrons from the radioactive process with this arrangement were preferentially thrown downwards towards the floor. From this it follows unambiguously that the process lacks that right-left symmetry, which one had earlier assumed. Thus, by means of this experiment it could be explained to a person, who did not know it – let us say an inhabitant of a distant stellar system – what we mean by right and left. In fact, it would be sufficient to ask him to arrange the experiment so as to make the preferential direction of the electrons point downwards. The current will then have the same direction as that in which he has to turn at the command «left face».

However – and this is a thing of the utmost importance for the incorporation of the new discoveries into our edifice of physical laws – the person on the distant planet will be able to follow our prescriptions only if he knows what we mean by the direction of an electric current. And to know this he must know that our atoms and his are made up of the same elementary particles. We know, however, that not only are there double sets of electrons – positive and negative – but that the same holds for protons and neutrons, the building stones of atomic nuclei. It is therefore possible that his atoms contrary to ours would consist of positive electrons and negative nuclei. If they did, he would judge the direction of the current opposite to what we would do, with the result that he would call right left and left right. In stating this we have tacitly made an assumption which is not quite confirmed as yet but which, as far as the experiments go, seems probable, namely that the results of all experiments performed with the opposite kind of elementary particles

would be just such as to re-establish the right-left symmetry. With other words, one should be able to regard the antiparticles not only as the electric opposites of the particles but also as their mirror images.

Professor Lee and Professor Yang. In the very incomplete sketch of your new work I have just made in Swedish, time has not allowed me to mention the many other beautiful contributions to theoretical physics made by each one of you, nor could I at all do justice to the enthusiasm your new achievement has aroused among physicists. Through your consistent and unprejudiced thinking you have been able to break a most puzzling deadlock in the field of elementary particle physics where now experimental and theoretical work is pouring forth as the result of your brilliant achievement.

It is with great satisfaction, therefore, that The Royal Swedish Academy of Sciences has decided to award you this year's Nobel Prize for Physics for your fundamental contributions to this branch of science.

On behalf of the Academy I wish to extend to you our heartiest congratulations and now ask you to receive from the hands of His Majesty the King the Nobel Prize for Physics for the year 1957.

CHEN NING YANG

The law of parity conservation and other symmetry laws of physics

Nobel Lecture, December 11, 1957

It is a pleasure and a great privilege to have this opportunity to discuss with you the question of parity conservation and other symmetry laws. We shall be concerned first with the general aspects of the role of the symmetry laws in physics; second, with the development that led to the disproof of parity conservation; and last, with a discussion of some other symmetry laws which physicists have learned through experience, but which do not yet together form an integral and conceptually simple pattern. The interesting and very exciting developments since parity conservation was disproved, will be covered by Dr. Lee in his lecture[1].

I

The existence of symmetry laws is in full accordance with our daily experience. The simplest of these symmetries, the isotropy and homogeneity of space, are concepts that date back to the early history of human thought. The invariance of physical laws under a coordinate transformation of uniform velocity, also known as the invariance under Galilean transformations, is a more sophisticated symmetry that was early recognized, and formed one of the corner-stones of Newtonian mechanics. Consequences of these symmetry principles were greatly exploited by physicists of the past centuries and gave rise to many important results. A good example in this direction is the theorem that in an isotropic solid there are only two elastic constants.

Another type of consequences of the symmetry laws relates to the conservation laws. It is common knowledge today that in general a symmetry principle (or equivalently an invariance principle) generates a conservation law. For example, the invariance of physical laws under space displacement has as a consequence the conservation of momentum, the invariance under space rotation has as a consequence the conservation of angular momentum. While the importance of these conservation laws was fully understood, their close relationship with the symmetry laws seemed not to have been clearly recognized until the beginning of the twentieth century[2]. (Cf. Fig. 1.)

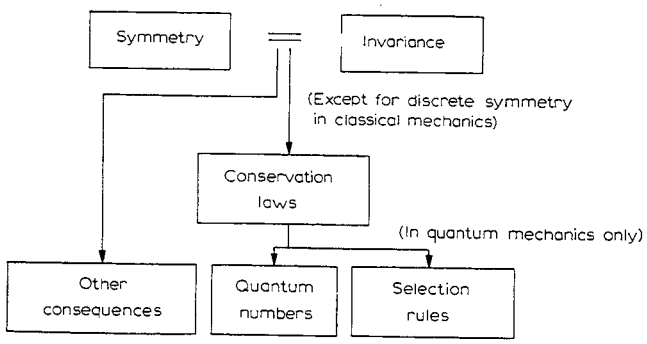


Fig. 1.

With the advent of special and general relativity, the symmetry laws gained new importance. Their connection with the dynamic laws of physics takes on a much more integrated and interdependent relationship than in classical mechanics, where logically the symmetry laws were only consequences of the dynamical laws that by chance possess the symmetries. Also in the relativity theories the realm of the symmetry laws was greatly enriched to include invariances that were by no means apparent from daily experience. Their validity rather was deduced from, or was later confirmed by complicated experimentation. Let me emphasize that the conceptual simplicity and intrinsic beauty of the symmetries that so evolve from complex experiments are for the physicists great sources of encouragement. One learns to hope that Nature possesses an order that one may aspire to comprehend.

It was, however, not until the development of quantum mechanics that the use of the symmetry principles began to permeate into the very language of physics. The quantum numbers that designate the states of a system are often identical with those that represent the symmetries of the system. It indeed is scarcely possible to overemphasize the role played by the symmetry principles in quantum mechanics. To quote two examples: The general structure of the Periodic Table is essentially a direct consequence of the isotropy of Coulomb's law. The existence of the antiparticles – namely the positron, the antiproton, and the antineutron – were theoretically anticipated as consequences of the symmetry of physical laws with respect to Lorentz transformations. In both cases Nature seems to take advantage of the simple mathematical representations of the symmetry laws. When one pauses to consider the elegance and the beautiful perfection of the mathematical reasoning involved and contrast it with the complex and far-reaching physical

consequences, a deep sense of respect for the power of the symmetry laws never fails to develop.

One of the symmetry principles, the symmetry between the left and the right, is as old as human civilization. The question whether Nature exhibits such symmetry was debated at length by philosophers of the past[3]. Of course, in daily life, left and right are quite distinct from each other. Our hearts, for example, are on our left sides. The language that people use both in the orient and the occident, carries even a connotation that right is good and left is evil. However, the laws of physics have always shown complete symmetry between the left and the right, the asymmetry in daily life being attributed to the accidental asymmetry of the environment, or initial conditions in organic life. To illustrate the point, we mention that if there existed a mirror-image man with his heart on his right side, his internal organs reversed compared to ours, and in fact his body molecules, for example sugar molecules, the mirror image of ours, and if he ate the mirror image of the food that we eat, then according to the laws of physics, he should function as well as we do.

The law of right-left symmetry was used in classical physics, but was not of any great practical importance there. One reason for this derives from the fact that right-left symmetry is a discrete symmetry, unlike rotational symmetry which is continuous. Whereas the continuous symmetries always lead to conservation laws in classical mechanics, a discrete symmetry does not. With the introduction of quantum mechanics, however, this difference between the discrete and continuous symmetries disappears. The law of right-left symmetry then leads also to a conservation law: the conservation of parity.

The discovery of this conservation law dates back to 1924 when Laporte[4] found that energy levels in complex atoms can be classified into « gestrichene » and « ungestrichene » types, or in more recent language, even and odd levels. In transitions between these levels during which one photon is emitted or absorbed, Laporte found that the level always changes from even to odd or *vice versa*. Anticipating later developments, we remark that the evenness or oddness of the levels was later referred to as the parity of the levels. Even levels are defined to have parity $+1$, odd levels parity -1. One also defines the photon emitted or absorbed in the usual atomic transitions to have odd parity. Laporte's rule can then be formulated as the statement that in an atomic transition with the emission of a photon, the parity of the initial state is equal to the total parity of the final state, i.e. the product of the parities of the final atomic state and the photon emitted. In other words, parity is conserved, or unchanged, in the transition.

In 1927 Wigner[5] took the critical and profound step to prove that the empirical rule of Laporte is a consequence of the reflection invariance, or right-left symmetry, of the electromagnetic forces in the atom. This fundamental idea was rapidly absorbed into the language of physics. Since right-left symmetry was unquestioned also in other interactions, the idea was further taken over into new domains as the subject matter of physics extended into nuclear reactions, β-decay, meson interactions, and strange-particle physics. One became accustomed to the idea of nuclear parities as well as atomic parities, and one discusses and measures the intrinsic parities of the mesons. Throughout these developments the concept of parity and the law of parity conservation proved to be extremely fruitful, and the success had in turn been taken as a support for the validity of right-left symmetry.

II

Against such a background the so-called ϑ—τ puzzle developed in the last few years. Before explaining the meaning of this puzzle it is best to go a little bit into a classification of the forces that act between subatomic particles, a classification which the physicists have learned through experience to use in the last 50 years. We list the four classes of interactions below. The strength of these interactions is indicated in the column on the right.

1. Nuclear Forces	1
2. Electromagnetic Forces	10^{-2}
3. Weak Forces (Decay Interactions)	10^{-13}
4. Gravitational Forces	10^{-38}

The strongest interactions are the nuclear interactions which include the forces that bind nuclei together and the interaction between the nuclei and the π mesons. It also includes the interactions that give rise to the observed strange-particle production. The second class of interactions are the electromagnetic interactions of which physicists know a great deal. In fact, the crowning achievement of the physicists of the 19th century was a detailed understanding of the electromagnetic forces. With the advent of quantum mechanics, this understanding of electromagnetic forces gives in principle an accurate, integral and detailed description of practically all the physical and chemical phenomena of our daily experience. The third class of forces, the weak interactions, was first discovered around the beginning of

this century in the β-radioactivity of nuclei, a phenomena which especially in the last 25 years has been extensively studied experimentally. With the discovery of π-μ, μ-e decays and μ capture it was noticed independently[6] by Klein, by Tiomno and Wheeler, and by Lee, Rosenbluth and me, that these interactions have roughly the same strengths as β-interactions. They are called weak interactions, and in the last few years their rank has been constantly added to through the discovery of many other weak interactions responsible for the decay of the strange particles. The consistent and striking pattern of their almost uniform strength remains today one of the most tantalizing phenomena – a topic which we shall come back to later. About the last class of forces, the gravitational forces, we need only mention that in atomic and nuclear interactions they are so weak as to be completely negligible in all the observations with existing techniques.

Now to return to the ϑ—τ puzzle. In 1953, Dalitz and Fabri[7] pointed out that in the decay of the ϑ and τ mesons

$$\vartheta \rightarrow \pi + \pi$$
$$\tau \rightarrow \pi + \pi + \pi$$

some information about the spins and parities of the τ and ϑ mesons can be obtained. The argument is very roughly as follows. It has previously been determined that the parity of a π meson is odd (i.e. $= -1$). Let us first neglect the effects due to the relative motion of the π mesons. To conserve parity in the decays, the ϑ meson must have the total parity, or in other words, the product parity, of two π mesons, which is even (i.e. $= +1$). Similarly, the τ meson must have the total parity of three π mesons, which is odd. Actually because of the relative motion of the π mesons the argument was not as simple and unambiguous as we just discussed. To render the argument conclusive and definitive it was necessary to study experimentally the momentum and angular distribution of the π mesons. Such studies were made in many laboratories, and by the spring of 1956 the accumulated experimental data seemed to unambiguously indicate, along the lines of reasoning discussed above, that ϑ and τ do not have the same parity, and consequently are not the same particle. This conclusion, however, was in marked contradiction with other experimental results which also became definite at about the same time. The contradiction was known as the ϑ—τ puzzle and was widely discussed. To recapture the atmosphere of that time allow me to quote a paragraph concerning the conclusion that ϑ and τ are not the same

particle from a report entitled «Present Knowledge about the New Particles» which I gave at the International Conference on Theoretical Physics[8] in Seattle, in September 1956.

«However it will not do to jump to hasty conclusions. This is because experimentally the K mesons (i.e. τ and ϑ) seem all to have the same masses and the same lifetimes. The masses are known to an accuracy of, say, from 2 to 10 electron masses, or a fraction of a percent, and the lifetimes are known to an accuracy of, say, 20 percent. Since particles which have different spin and parity values, and which have strong interactions with the nucleons and pions, are not expected to have identical masses and lifetimes, one is forced to keep the question open whether the inference mentioned above that the τ^+ and ϑ^+ are not the same particle is conclusive. *Parenthetically, I might add that the inference would certainly have been regarded as conclusive, and in fact more well-founded than many inferences in physics, had it not been for the anomaly of mass and lifetime degeneracies.*»

The situation that the physicist found himself in at that time has been likened to a man in a dark room groping for an outlet. He is aware of the fact that in some direction there must be a door which would lead him out of his predicament. But in which direction?

That direction turned out to lie in the faultiness of the law of parity conservation for the weak interactions. But to uproot an accepted concept one must first demonstrate why the previous evidence in its favor were insufficient. Dr. Lee and I[9] examined this question in detail, and in May 1956 we came to the following conclusions: (*A*) Past experiments on the weak interactions had actually no bearing on the question of parity conservation. (*B*) In the strong interactions, i.e. interactions of classes 1 and 2 discussed above, there were indeed many experiments that established parity conservation to a high degree of accuracy, but not to a sufficiently high degree to be able to reveal the effects of a lack of parity conservation in the weak interactions.

The fact that parity conservation in the weak interactions was believed for so long without experimental support was very startling. But what was more startling was the prospect that a space-time symmetry law which the physicists have learned so well may be violated. This prospect did not appeal to us. Rather we were, so to speak, driven to it through frustration with the various other efforts at understanding the ϑ—τ puzzle that had been made[10].

As we shall mention later there is known in physics a conservation law – the conservation of isotopic spin – that holds for interactions of class 1 but breaks down when weaker interactions are introduced. Such a possibility of

an approximate symmetry law was, however, not expected of the symmetries related to space and time. In fact one is tempted to speculate, now that parity conservation is found to be violated in the weak interactions, whether in the description of such phenomena the usual concept of space and time is adequate. At the end of our discussion we shall have the occasion to come back to a closely related topic.

Why was it so that among the multitude of experiments on β-decay, the most exhaustively studied of all the weak interactions, there was no information on the conservation of parity in the weak interactions? The answer derives from a combination of two reasons. First, the fact that the neutrino does not have a measurable mass introduces an ambiguity that rules out[11] indirect information on parity conservation from such simple experiments as the spectrum of β-decay. Second, to study directly parity conservation in β-decay it is not enough to discuss nuclear parities, as one had always done. One must study parity conservation of the *whole* decay process. In other words, one must design an experiment that tests right-left symmetry in the decay. Such experiments were not done before.

Once these points were understood it was easy to point out what were the experiments that would unambiguously test the previously untested assumption of parity conservation in the weak interactions. Dr. Lee and I proposed[9] in the summer of 1956 a number of these tests concerning β-decay, π-μ, μ-e and strange-particle decays. The basic principles involved in these experiments are all the same: *One constructs two sets of experimental arrangements which are mirror images of each other, and which contain weak interactions. One then examines whether the two arrangements always give the same results in terms of the readings of their meters* (or counters). If the results are not the same, one would have an unequivocal proof that right-left symmetry, as we usually understand it, breaks down. The idea is illustrated in Fig. 2 which shows the experiment proposed to test parity conservation in β-decay.

This experiment was first performed in the latter half of 1956 and finished early this year by Wu, Ambler, Hayward, Hoppes, and Hudson[12]. The actual experimental setup was very involved, because to eliminate disturbing outside influences the experiment had to be done at very low temperatures. The technique of combining β-decay measurement with low temperature apparatus was unknown before and constituted a major difficulty which was successfully solved by these authors. To their courage and their skill, physicists owe the exciting and clarifying developments concerning parity conservation in the past year.

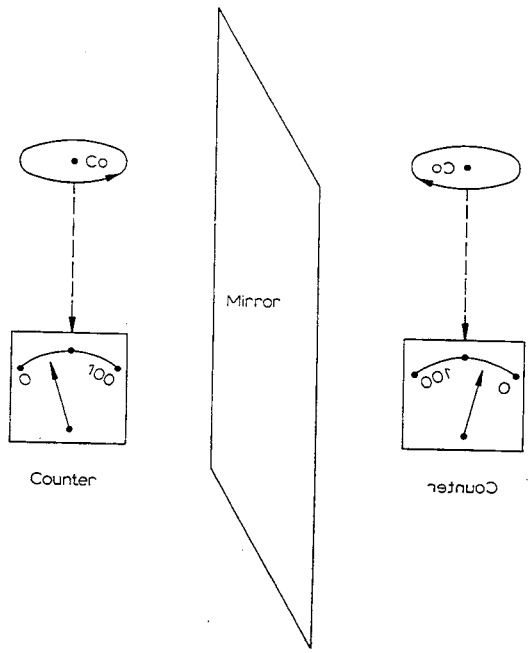


Fig. 2.

The results of Drs. Wu, Ambler, and their collaborators was that there is a very large difference in the readings of the two meters of Fig. 2. Since the behavior of the other parts of their apparatus observes right-left symmetry, the asymmetry that was found must be attributed to the β-decay of cobalt. Very rapidly after these results were made known, many experiments were performed which further demonstrated the violation of parity conservation in various weak interactions. In his lecture[1] Dr. Lee will discuss these interesting and important developments.

III

The breakdown of parity conservation brings into focus a number of questions concerning symmetry laws in physics which we shall now briefly discuss in general terms:

(*A*) As Dr. Lee[1] will discuss, the experiment of Wu, Ambler, and their collaborators also proves[13,14] that charge conjugation invariance[15] is violated for β-decay. Another symmetry called time reversal invariance[16] is at the present moment still being experimentally studied for the weak interactions.

The three discrete invariances – reflection invariance, charge conjugation invariance, and time reversal invariance – are connected by an important theorem[17] called the CPT theorem. Through the use of this theorem one can prove[13] a number of general results concerning the experimental manifestations of the possible violations of the three symmetries in the weak interactions.

Of particular interest is the *possibility* that time reversal invariance in the weak interactions may turn out to be intact. If this is the case, it follows from the CPT theorem that although parity conservation breaks down, right-left symmetry will still hold if[18] one switches all particles into antiparticles in taking a mirror image. In terms of Fig. 2 this means that if one changes *all* the matter that composes the apparatus at the right into anti-matter, the meter reading would become the same for the two sides if time reversal invariance holds. It is important to notice that in the usual definition of reflection, the electric field is a vector and the magnetic field a pseudovector while in this changed definition their transformation properties are switched. The transformation properties of the electric charge and the magnetic charge are also interchanged. It would be interesting to speculate on the possible relationship between the nonconservation of parity and the symmetrical or unsymmetrical role played by the electric and magnetic fields.

The question of the validity of the continuous space time symmetry laws has been discussed to some extent in the past year. There is good evidence that these symmetry laws do not break down in the weak interactions.

(*B*) Another symmetry law that has been widely discussed is that giving rise to the conservation of isotopic spin[19]. In recent years the use of this symmetry law has produced a remarkable empirical order among the phenomena concerning the strange particles[20]. It is however certainly the least understood of all the symmetry laws. Unlike Lorentz invariance or reflection invariance, it is not a «geometrical» symmetry law relating to space time invariance properties. Unlike charge conjugation invariance[21] it does not seem to originate from the algebraic property of the complex numbers that occurs in quantum mechanics. In these respects it resembles the conservation laws of charge and heavy particles. These latter laws, however, are exact while the conservation of isotopic spin is violated upon the introduction of electromagnetic interactions and weak interactions. An understanding of the origin of the conservation of isotopic spin and how to integrate it with the other symmetry laws is undoubtedly one of the outstanding problems in high-energy physics today.

(C) We have mentioned before that all the different varieties of weak interactions share the property of having very closely identical strengths. The experimental work on parity nonconservation in the past year reveals that they very likely also share the property of not respecting parity conservation and charge conjugation invariance. They therefore serve to differentiate between right and left once one fixes one's definition of matter *vs.* anti-matter. One could also use the weak interactions to differentiate between matter and anti-matter once one chooses a definition of right *vs.* left. If time reversal invariance is violated, the weak interactions may even serve to differentiate simultaneously right from left, and matter from anti-matter. One senses herein that maybe the origin of the weak interactions is intimately tied in with the question of the differentiability of left from right, and of matter from anti-matter.

1. T. D. Lee, *Nobel Lecture*, this volume, p. 406.
2. For references to these developments see E. P. Wigner, *Proc. Am. Phil. Soc.*, 93 (1949) 521.
3. Cf. the interesting discussion on bilateral symmetry by H. Weyl, *Symmetry*, Princeton University Press, 1952.
4. O. Laporte, *Z. Physik*, 23 (1924) 135.
5. E. P. Wigner, *Z. Physik*, 43 (1927) 624.
6. O. Klein, *Nature*, 161 (1948) 897; J. Tiomno and J. A. Wheeler, *Rev. Mod. Phys.*, 21 (1949) 144; T. D. Lee, M. Rosenbluth, and C. N. Yang, *Phys. Rev.*, 75 (1949) 905.
7. R. Dalitz, *Phil. Mag.*, 44 (1953) 1068; E. Fabri, *Nuovo Cimento*, 11 (1954) 479.
8. C. N. Yang, *Rev. Mod. Phys.*, 29 (1957) 231.
9. T. D. Lee and C. N. Yang, *Phys. Rev.*, 104 (1956) 254.
10. T. D. Lee and J. Orear, *Phys. Rev.*, 100 (1955) 932; T. D. Lee and C. N. Yang, *Phys. Rev.*, 102 (1956) 290; M. Gell-Mann, (unpublished); R. Weinstein, (private communication); a general discussion of these ideas can be found in the *Proceedings of the Rochester Conference*, April 1956, Session VIII, Interscience, New York, 1957.
11. C. N. Yang and J. Tiomno, *Phys. Rev.*, 79 (1950) 495.
12. C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Phys. Rev.*, 105 (1957) 1413.
13. T. D. Lee, R. Oehme, and C. N. Yang, *Phys. Rev.*, 106 (1957) 340.
14. B. L. Ioffe, L. B. Okun, and A. P. Rudik, *J.E.T.P.* (*U.S.S.R.*), 32 (1957) 396. English translation in *Soviet Phys. JETP*, 5 (1957) 328.
15. Charge conjugation invariance is very intimately tied with the hole theory interpretation of Dirac's equation. The development of the latter originated with P. A. M. Dirac, *Proc. Roy. Soc. London*, A126 (1930) 360; J. R. Oppenheimer, *Phys. Rev.*,

35 (1930) 562 and H. Weyl, *Gruppentheorie und Quantenmechanik*, 2nd ed., 1931, p. 234. An account of these developments is found in P. A. M. Dirac, *Proc. Roy. Soc. London*, A133 (1931) 60. Detailed formalism and application of charge conjugation invariance started with H. A. Kramers, *Proc. Acad. Sci. Amsterdam*, 40 (1937) 814 and W. Furry, *Phys. Rev.*, 51 (1937) 125.

16. E. P. Wigner, *Nachr. Akad. Wiss. Goettingen, Math.-Physik.*, 1932, p. 546. This paper explains in terms of time reversal invariance the earlier work of H. Kramers, *Proc. Acad. Sci. Amsterdam*, 33 (1930) 959.
17. J. Schwinger, *Phys. Rev.*, 91 (1953) 720, 723; G. Lüders, *Kgl. Danske Videnskab. Selskab., Mat.-Fys. Medd.*, 28, No. 5 (1954); W. Pauli's article in *Niels Bohr and the Development of Physics*, Pergamon Press, London, 1955. See also Ref. 21.
18. This possibility was discussed by T. D. Lee and C. N. Yang and reported by C. N. Yang at the International Conference on Theoretical Physics in Seattle in September 1956. (See Ref. 8.) Its relation with the CPT theorem was also reported in the same conference in one of the discussion sessions. The speculation was later published in T. D. Lee and C. N. Yang, *Phys. Rev.*, 105 (1957) 1671. Independently the possibility has been advanced as the correct one by L. Landau, *J.E.T.P. (U.S.S.R.)*, 32 (1957) 405. An English translation of Landau's article appeared in *Soviet Phys. JETP*, 5 (1957) 336.
19. The concept of a total isotopic spin quantum number was first discussed by B. Cassen and E. U. Condon, *Phys. Rev.*, 50 (1936) 846 and E. P. Wigner, *Phys. Rev.*, 51 (1937) 106. The physical basis derived from the equivalence of p-p and n-p forces, pointed out by G. Breit, E. U. Condon, and R. D. Present, *Phys. Rev.*, 50 (1936) 825. The isotopic spin was introduced earlier as a formal mathematical parameter by W. Heisenberg, *Z. Physik*, 77 (1932) 1.
20. A. Pais, *Phys. Rev.*, 86 (1952) 663, introduced the idea of associated production of strange particles. An explanation of this phenomenon in terms of isotopic spin conservation was pointed out by M. Gell-Mann, *Phys. Rev.*, 92 (1953) 833 and by K. Nishijima, *Progr. Theoret. Phys. (Kyoto)*, 12 (1954) 107. These latter authors also showed that isotopic spin conservation leads to a convenient quantum number called strangeness.
21. R. Jost, *Helv. Phys. Acta*, 30 (1957) 409.

Biography

Chen Ning Yang was born on September 22, 1922, in Hofei, Anwhei, China, the first of five children of Ke Chuan Yang and Meng Hwa Loh Yang. He is also known as Frank or Franklin.

Yang was brought up in the peaceful and academically inclined atmosphere of the campus of Tsinghua University, just outside of Peiping, China, where his father was a Professor of Mathematics. He received his college education at the National Southwest Associated University in Kunming, China, and completed his B.Sc. degree there in 1942. His M.Sc. degree was received in 1944 from Tsinghua University, which had moved to Kunming during the Sino-Japanese War (1937–1945). He went to the U.S.A. at the end of the war on a Tsinghua University Fellowship, and entered the University of Chicago in January 1946. At Chicago he came under the strong influence of Professor E. Fermi. After receiving his Ph.D. degree in 1948, Yang served for a year at the University of Chicago as an Instructor. He has been associated with the Institute for Advanced Study, Princeton, New Jersey, U.S.A., since 1949, where he became a Professor in 1955.

Yang has worked on various subjects in physics, but has his chief interest in two fields: statistical mechanics and symmetry principles. His B.Sc. thesis: « Group Theory and Molecular Spectra », written under the guidance of Professor Ta-You Wu, his M.Sc. thesis: « Contributions to the Statistical Theory of Order-Disorder Transformations », written under the guidance of Professor J. S. Wang, and his Ph.D. thesis: « On the Angular Distribution in Nuclear Reactions and Coincidence Measurements », written under the guidance of Professor E. Teller, were instrumental in introducing him to these fields.

Dr. Yang is a prolific author, his numerous articles appearing in the *Bulletin of the American Mathematical Society*, *The Physical Review*, *Reviews of Modern Physics*, and the *Chinese Journal of Physics*.

Professor Yang has been elected Fellow of the American Physical Society and the Academia Sinica, and honoured with the Albert Einstein Commemorative Award (1957). The U.S. Junior Chamber of Commerce named

him one of the outstanding young men of 1957. He was also awarded an honorary doctorate of the Princeton University, N.J. (1958).

In 1950 Yang married Chih Li Tu and is now the father of three children: Franklin, born 1951; Gilbert, born 1958; and Eulee, born 1961.

Dr. Yang is a quiet, modest, and affable physicist; he met his wife Chih Li Tu while teaching mathematics at her high school in China. He is a hard worker allowing himself very little leisure time.

TSUNG DAO LEE

Weak interactions and nonconservation of parity

Nobel Lecture, December 11, 1957

In the previous talk Professor Yang[1] has outlined to you the position of our understanding concerning the various symmetry principles in physics prior to the end of last year. Since then, in the short period of one year, the proper roles of these principles in various physical processes have been greatly clarified. This remarkably rapid development is made possible only through the efforts and ingenuity of many physicists in various laboratories all over the world. To have a proper perspective and understanding of these new experimental results it may be desirable to review very briefly our knowledge about elementary particles and their interactions.

I

The family of elementary particles that we know today consists of numerous members. Each member is characterized, among other properties, by its mass, charge, and spin. These members are separated into two main groups: the «heavy particle» group and the «light particle» group. The well-known examples of the heavy particles are protons and neutrons; those of the light particles are photons and electrons. Apart from the obvious implication that a heavy particle is heavier than a light particle, this classification stems from the observation that a single heavy particle cannot disintegrate into light particles even if such disintegration should be compatible with the conservation laws of charge, energy, momentum, and angular momentum. This fact is more precisely formulated as the «law of conservation of heavy particles» which states that if to each heavy particle we assign a heavy particle number $+1$, to each anti-heavy particle a heavy particle number -1, and to each light particle a corresponding number 0, then in all known physical processes the algebraic sum of the heavy particle numbers is absolutely conserved. One of the simplest evidences of the validity of this law is the fact that ourselves, or our galaxy, have not disintegrated into radiation and other light particles.

Fig. 1 shows all the known heavy particles (and anti-heavy particles). All

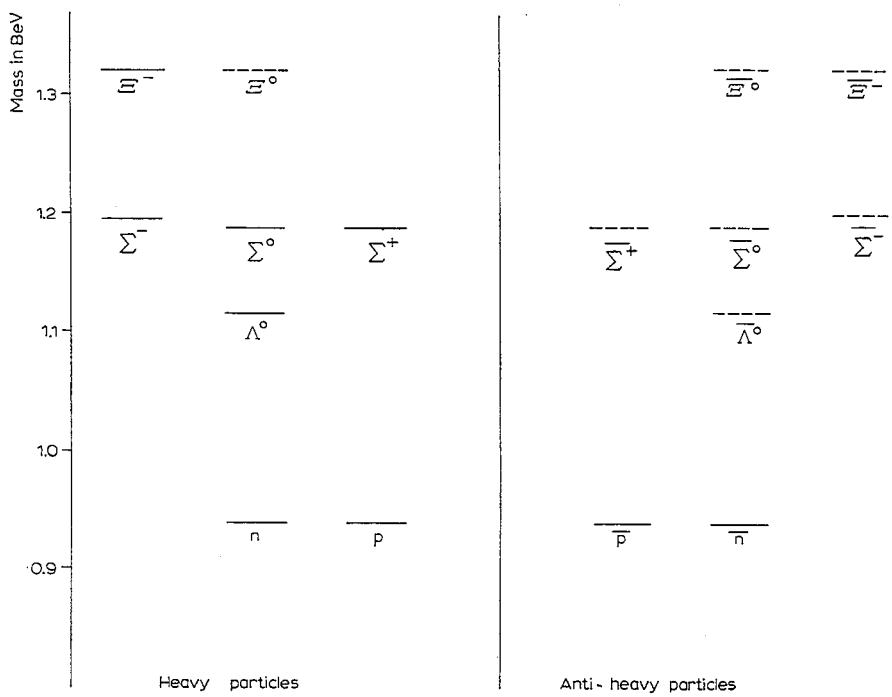


Fig. 1.

heavy particles except the nucleons are called hyperons and are labelled by capital Greek letters. The solid lines represent particles that are expected to exist from general theoretical arguments. All known heavy particles have half-integral spins. Fig. 2 shows all the known light particles. Among these, the $e^{\pm}$, $\mu^{\pm}$ and ν, $\bar{\nu}$ have half-integral spins. They are called leptons. The rest, photons, pions and K mesons, have integral spins.

The interactions (not including the gravitational forces) between these particles can be classified into three distinct groups:

1. *Strong Interactions.* This group is responsible for the production and the scattering of nucleons, pions, hyperons (i.e. Λ°, Σ^{-}, etc.) and K mesons. It is characterized by a coupling constant $f^2/\hbar c \cong 1$.

2. *Electromagnetic Interactions.* The electromagnetic coupling constant is $(e^2/\hbar c) = (1/137)$.

3. *Weak Interactions.* This group includes all known non-electromagnetic decay interactions of these elementary particles and the recently observed absorption process of neutrinoes by nucleons[2]. These interactions are characterized by coupling constants $g^2/\hbar c \cong 10^{-14}$.

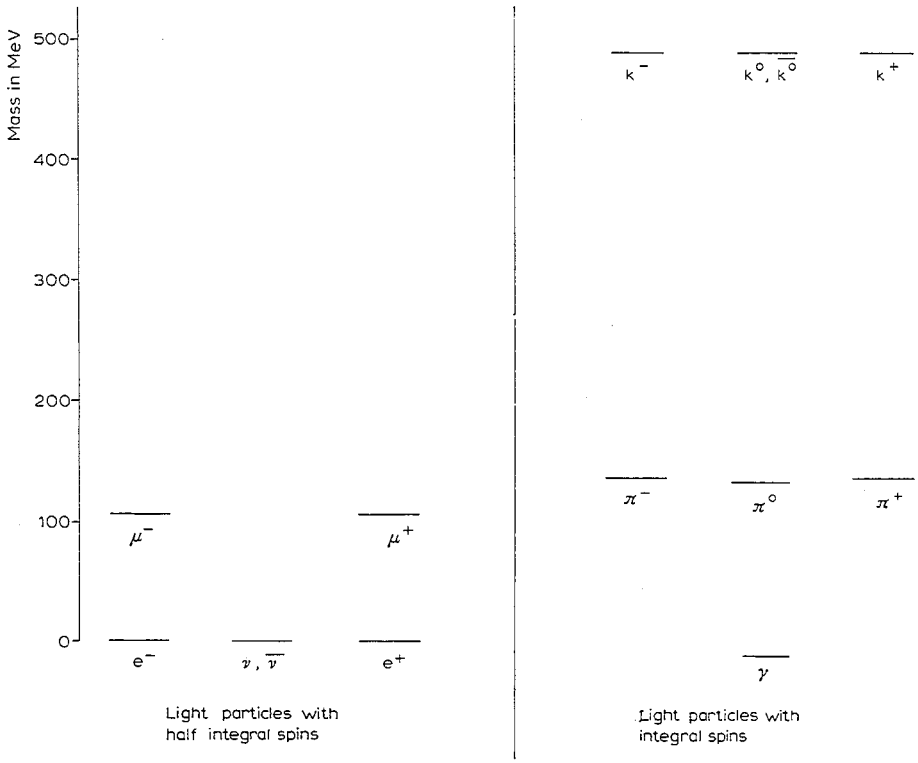


Fig. 2.

The law of conservation of parity is valid for both the strong and the electromagnetic interactions but is not valid for the weak interactions. Today's discussions will be mainly on the recently observed effects of nonconservation of parity in the various weak interactions.

II

The weak interactions cover a large variety of reactions. At present there are about 20 known phenomenologically independent reactions ranging from the decay of various hyperons to the decay of light particles. Within the last year, many critical experiments have been performed to test the validity of the law of conservation of parity in these reactions. We shall first summarize the experimental results together with their direct theoretical implications. Next, we shall discuss some further possible consequences and theoretical considerations.

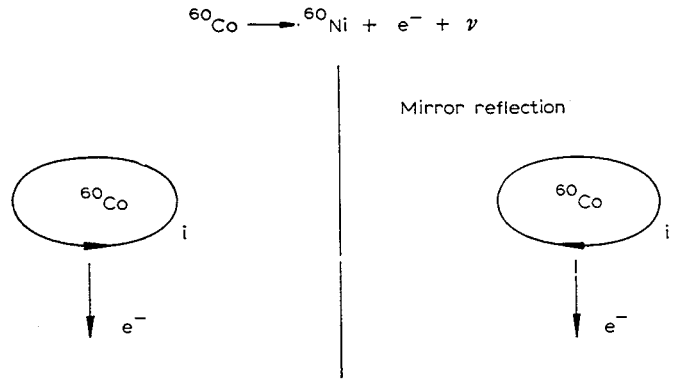


Fig. 3.

(1) *β-decay*

The first experiment that conclusively established the nonconservation of parity was that on β-angular distribution from polarized ^{60}Co nuclei[3] (see Fig. 3). The ^{60}Co nuclei are polarized by a magnetic field at very low temperatures. Indeed in this experiment, the circular direction of the electric current in the solenoid that produces the polarizing magnetic field together with the preferential direction of the β-ray emitted, differentiates in a most direct way a right-handed system from a left-handed system. Thus the nonconservation of parity or the non-invariance under a mirror reflection can be established without reference to any theory.

Furthermore from the large amount of angular asymmetry observed it can also be established[4] that the β-decay interaction is not invariant under a charge conjugation operation. That this can be concluded without performing the extremely difficult (in fact, almost impossible) experiment using anti-^{60}Co is based on certain theoretical deductions under the general framework of local field theory. In the following we shall try to sketch this type of reasoning[5].

Let us consider the β-decay process, say

$$n \rightarrow p + e^- + \nu \tag{1}$$

in which each particle is described by a quantized wave equation. In particular the neutrino is described by the Dirac equation[6]

$$\sum_{\mu=1}^{4} \gamma_\mu \frac{\partial}{\partial x_\mu} \psi_\nu = 0 \tag{2}$$

where $\gamma_1, \gamma_2, \gamma_3, \gamma_4$, are the four (4×4) anti-commuting Dirac matrices and $x_1, x_2, x_3, x_4 = ict$ are the four space-time coordinates. For each given momentum there exists two spin states for the neutrino and two spin states for the anti-neutrino. These may be denoted by ν_R, ν_L, $\bar{\nu}_R$, $\bar{\nu}_L$. If we define the helicity H to be

$$H \equiv \vec{\sigma} \cdot \hat{p} \tag{3}$$

with $\vec{\sigma}$ as the spin operator and $\hat{p}$ the unit vector along the momentum direction, then these four states have, respectively, helicities equal to $+1$, -1, -1 and $+1$ (Fig. 4). Mathematically, this decomposition of states corresponds to a separation of ψ_ν into a right-handed part ψ_R and a left-handed part ψ_L with

$$\psi_\nu = \psi_R + \psi_L \tag{4}$$

where

$$\psi_R = \tfrac{1}{2}(1 - \gamma_5)\,\psi_\nu \tag{5}$$

$$\psi_L = \tfrac{1}{2}(1 + \gamma_5)\,\psi_\nu \tag{6}$$

and

$$\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4$$

It is easy to see that both ψ_R and ψ_L separately satisfy the Dirac equation [Eq. (2)]. With this decomposition the β process of a nucleus A can be represented schematically as

$$\mathrm{A} \to \mathrm{B} + e^- + \begin{cases} C_i^R \nu_R & (H = +1) \quad (7) \\ C_i^L \nu_L & (H = -1) \quad (8) \end{cases}$$

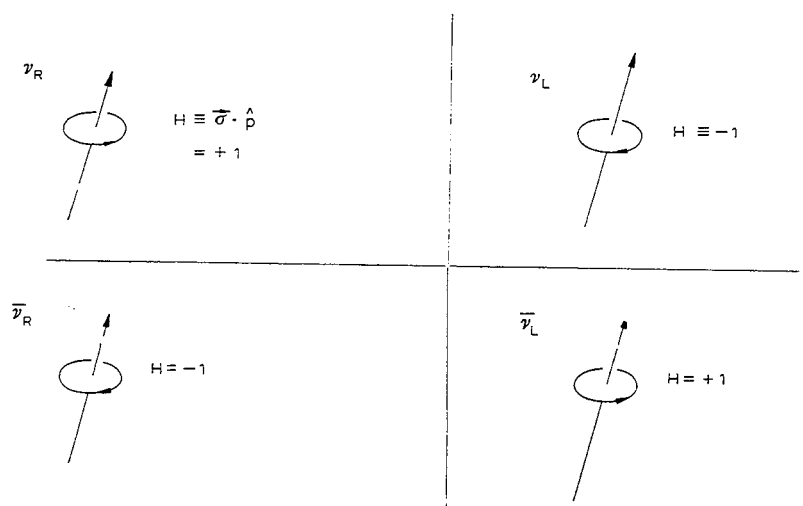


Fig. 4.

with C_i^R and C_i^L as the various probability amplitudes for emission of ν_R and ν_L respectively. The suffix i represents the various possible channels for such emissions. If the theory is invariant under proper Lorentz transformation, then there are five such channels: namely scalar S, tensor T, vector V, pseudo-scalar P and axial-vector term A. According to the general rules of quantum field theory with any interaction term representing the decay of a particle, there exists a corresponding hermitian conjugate term which represents decay of the antiparticle. Thus, the decay of the anti-nucleus $\bar{A}$ can be schematically represented by

$$\bar{A} \rightarrow \bar{B} + e^+ + \begin{cases} C_i^{R*}\,\bar{\nu}_R & (H = -1) \quad (7') \\ C_i^{L*}\,\bar{\nu}_L & (H = +1) \quad (8') \end{cases}$$

with C_i^{R*} and C_i^{L*} as the corresponding amplitudes for emission of $\bar{\nu}_R$ and $\bar{\nu}_L$. Under the charge conjugation operator we change a particle to its anti-particle but we do not change its spatial or spin wave functions. Consequently it must have the same helicity. Thus, if the β-decay process is invariant under the charge conjugation operator, then we should expect process (7) to proceed with the same amplitude as process (8'). The condition for invariance under charge conjugation is, then

$$C_i^R = C_i^{L*} \qquad (9)$$

for all i = S, T, V, P, A.

In the decay of ^{60}Co, because there is a difference of spin values between ^{60}Co and ^{60}Ni, only the terms i = T and i = A contribute. From the large angular-asymmetry observed it can be safely concluded that for both i = T, A

$$|C_i^R| \neq |C_i^L|$$

which contradicts Eq. (9) and proves the non-invariance of β-interaction under charge conjugation. For illustration purposes, we assume in the above the neutrino to be described by a 4-component theory and further we assume that in the β-decay process only neutrino is emitted. Actually the same conclusion concerning the non-invariance property under charge conjugation can be obtained even if the neutrino should be described by a, say, 8-component theory, or, if in addition to neutrino, anti-neutrino may also be emitted.

Recently many more experiments[7] have been performed on the longitudinal polarization of electrons and positrons, the β-γ correlation together with the circular polarization of the γ radiation and the β angular distribution with various polarized nuclei other than ^{60}Co. The results of all these experiments confirm the main conclusions of the first ^{60}Co experiment, that both the parity operator and the charge conjugation operator are not conserved in β-decay processes.

Another interesting question is whether the β-decay interaction is invariant under the product operation of (charge conjugation $\times$ mirror reflection). Under such an operation we should compare the decay of A with that of $\bar{A}$ but with opposite helicities. Thus if β-decay is invariant under the joint operation of (charge conjugation $\times$ minor reflection) we should expect process (7) to proceed with the same amplitude as process (7′) and similarly for processes (8) and (8′). The corresponding conditions are then

$$\begin{aligned} & C_i^R = C_i^{R*} \\ \text{and} \quad & C_i^L = C_i^{L*} \end{aligned} \tag{10}$$

Although experiments have been performed to test the validity of these conditions, at present, these experiments have not reached a conclusive stage and we still do not know the answer to this important question.

(2) π-μ-e *decay*

The $\pi^{\pm}$ meson decays into a $\mu^{\pm}$ meson and a neutrino. The $\mu^{\pm}$ meson, in turn, decays into an $e^{\pm}$ and two neutrinoes (or anti-neutrinoes). If parity is not conserved in π-decay, the μ meson emitted could be longitudinally po-

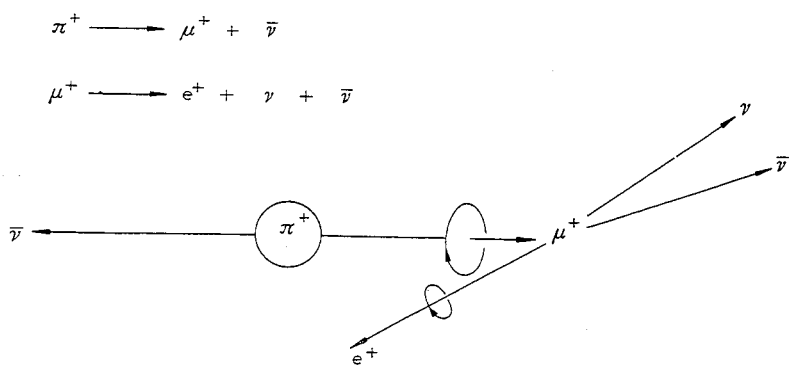


Fig. 5.

larized. If in the subsequent μ-decay parity is also not conserved, the electron (or positron) emitted from such a μ meson at rest would in general exhibit a forward and backward angular asymmetry with respect to the polarization of μ meson (Fig. 5). Consequently in the π-μ-e decay sequence we may observe an angular correlation between the momentum of $\mu^{\pm}$ meson measured in the rest system of π meson and the momentum of $e^{\pm}$ measured in the rest system of $\mu^{\pm}$. If this angular correlation shows a forward backward asymmetry, then parity must be nonconserved in both π-decay and μ-decay. The experimental results[8] on these angular correlations appeared within a few days after the results on β-decay were known. These results showed conclusively that not only parity is not conserved but the charge conjugation operator is also not conserved in π-decay as well as in μ-decay.

Later, direct measurements[9] on the longitudinal polarization of the positron from μ^{+}-decay was done establishing the same conclusion concerning μ-decay.

(3) K-μ-e *decay*

In this case we have instead of the π meson the heavier K meson which decays into a μ meson and a neutrino (Fig. 6). Experiment[10] on the angular correlation between the μ^{+} momentum from the decay of K^{+} meson and the positron momentum from the $\mu^{\pm}$-decay establishes that in K-decay the parity as well as the charge conjugation operator is not conserved.

(4) Λ°*-decay*

The Λ° particle can be produced by colliding an energetic π^{-} on proton. The

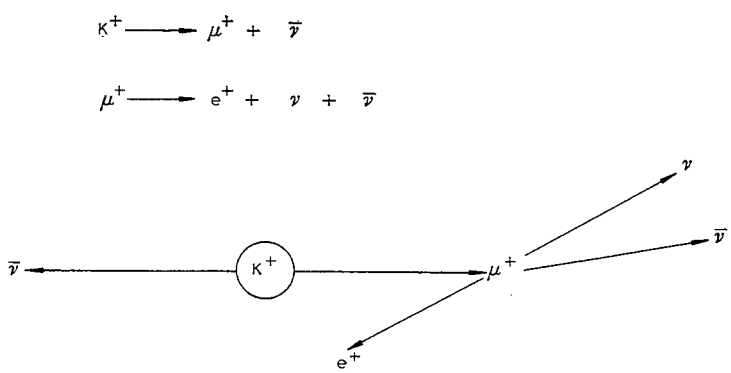


Fig. 6.

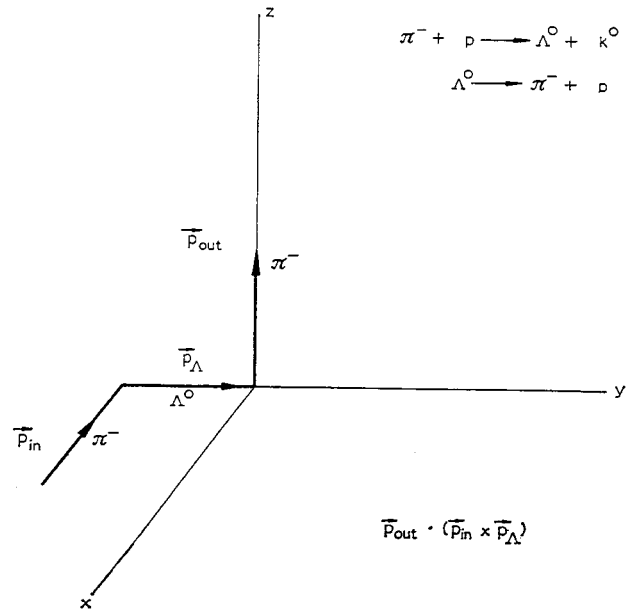


Fig. 7.

Λ° subsequently decays into a proton plus a π^- (Fig. 7). The observation of an assymetrical distribution with respect to the sign of the product $\vec{p}_{out}$ $(\vec{p}_{in} \times \vec{p}_\Lambda)$ formed from the momentum of the incoming pion $\vec{p}_{in}$, the momentum of the lambda particle, $\vec{p}_\Lambda$, and that of the decay pion $\vec{p}_{out}$ would constitute an unequivocal proof that parity is not conserved in this decay. Recent experiments[11] on these reactions demonstrates that in these reactions there is indeed such an angular correlation between $\vec{p}_{out}$ and $(\vec{p}_{in} \times \vec{p}_\Lambda)$. Furthermore, from the amount of the large up-down asymmetry it can be concluded that the Λ°-decay interaction is also not invariant under the charge conjugation operation.

From all these results it appears that the property of nonconservation of parity in the various weak interactions and the noninvariance property of these interactions under charge conjugation are well established. In connection with these properties we find an entirely new and rich domain of natural phenomena which, in turn, gives us new tools to probe further into the structure of our physical world. These interactions offer us natural ways to polarize and to analyze the spins of various elementary particles. Thus, for example, the magnetic moment of the μ meson can now be measured to an extremely high degree of accuracy[12] which, otherwise, would be unattain-

able; the spins of some hyperons now may perhaps be determined[13] unambiguously through the observed angular asymmetries in their decays; new aspects of the electromagnetic fields of various gas, liquid and solid materials can now be studied by using these unstable, polarized particles. However, perhaps the most significant consequences are the opening of new possibilities and the re-examination of our old concepts concerning the structure of elementary particles. We shall next discuss two such considerations – the two-component theory of neutrino, and the possible existence of a law of conservation of leptons.

III

Before the recent developments on nonconservation of parity, it was customary to describe the neutrino by a four-component theory in which, as we mentioned before, to each definite momentum there are the two spin states of the neutrino ν_R and ν_L, plus the two spin states of the antineutrino $\bar{\nu}_R$ and $\bar{\nu}_L$. In the two-component theory, however, we assume two of these states, say, ν_L and $\bar{\nu}_L$ simply do not exist in nature. The spin of the neutrino is then always parallel to its momentum while the spin of the antineutrino is always antiparallel to its momentum. Thus in the two-component theory we have only half of the degrees of freedom as in the four-component theory. Graphically we may represent the spin and the velocity of the neutrino by the spiral motion of a right-handed screw and that of the antineutrino by the motion of a left-handed screw (Fig. 8).

The possibility of a two-component relativistic theory of a spin $\frac{1}{2}$ particle was first discussed by H. Weyl[14] as early as 1929. However, in the past, because parity is not manifestly conserved in the Weyl formalism, it was always rejected[15]. With the recent discoveries such an objection becomes completely invalid[16].

To appreciate the simplicity of this two-component theory in the present situation it is best if we assume further the existence of a conservation law for leptons[17]. This law is in close analogy with the corresponding conservation law for the heavy particles. We assign to each lepton a leptonic number l equal to $+1$ or -1 and to any other particle the leptonic number zero. The leptonic number for a lepton must be the negative of that for its antiparticle. The law of conservation of leptons then states that «in all physical processes the algebraic sum of leptonic numbers must be conserved».

Some simple consequences follow immediately if we assume that this law is valid and that the neutrino is described by the two-component theory.
(1) The mass of the neutrino and the antineutrino must be zero. This is true

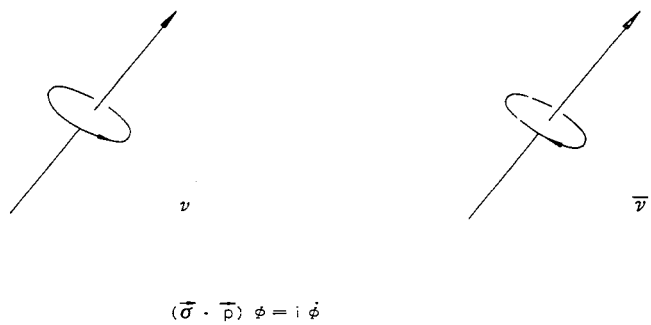


Fig. 8.

for the physical mass even with the inclusion of all interactions. To see this let us consider a neutrino moving with a finite momentum. From the two-component theory the spin of this neutrino must be parallel to its momentum. Suppose now it has a non-vanishing physical mass. Then, we can always send an observer travelling along the same direction as the neutrino but with a velocity faster than that of the neutrino. From this observer's point of view this « neutrino » now becomes a particle with the spin along its original direction but the direction of momentum reversed; i.e. it becomes an « antineutrino ». However since the leptonic number for neutrino is different from that of the antineutrino, these two particles cannot be transformed into each other by a Lorentz transformation. Consequently, the physical mass of a neutrino must be zero.
(2) The theory is not invariant under the parity operator P which by definition inverts all spatial coordinates but does not change a particle into its antiparticle state. Under such an operation one inverts the momentum of a particle but not its spin direction. Since in this theory these two are always parallel for a neutrino, the parity operator P applied to a neutrino state leads to a non-existing state. Consequently the theory is not invariant under the parity operation.
(3) Similarly one can show the theory is not invariant under the charge conjugation operation which changes a particle into its antiparticle but not its spin direction or its momentum.

To test the complete validity of the conservation law of leptons and the

two-component theory we have to investigate in detail all the neutrino processes. For example in β-decay we must have either

$$\begin{aligned} & n \to p + e^- + \nu \quad (H_\nu = +1) \\ \text{or} \quad & n \to p + e^- + \bar{\nu} \quad (H_{\bar{\nu}} = -1) \end{aligned}$$

This can be determined by measuring the spin and the momentum of the neutral lepton; i.e. to see whether it is a neutrino (right-handed helicity) or an antineutrino (left-handed helicity). Through the law of conservation of angular momentum, measurements on polarizations and angular distributions of the nucleus and the electrons can lead to determination of the spin states of the neutrino. Similarly, through the recoil momentum measurements we can find out information about the linear momentum of the neutrino. In the same way we can use not only β-decay but π-decay, μ-decay and K-decay to test the validity of either the two-component theory or the law of conservation of leptons. At present, these measurements have not yet reached a definitive stage. Much of our future may depend on the results of these experiments.

IV

The progress of science has always been the result of a close interplay between our concepts of the universe and our observations on nature. The former can only evolve out of the latter and yet the latter is also conditioned greatly by the former. Thus in our exploration of nature, the interplay between our concepts and our observations may sometimes lead to totally unexpected aspects among already familiar phenomena. As in the present case, these hidden properties are usually revealed only through a fundamental change in our basic concept concerning the principles that underly natural phenomena. While all this is well-known, it is nevertheless an extremely rich and memorable experience to be able to watch at close proximity in a single instance the mutual influence and the subsequent growth of these two factors – the concept and the observation. It is, indeed, a privilege that I am able to tell you part of this experience in the recent developments concerning the nonconservation of parity and the weak interactions.

1. C. N. Yang, *Nobel Lecture*, this volume, p. 393.
2. C. L. Cowan, Jr., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, *Science*, 124 (1956) 103.
3. C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Phys. Rev.*, 105 (1957) 1413.
4. T. D. Lee, R. Oehme, and C. N. Yang, *Phys. Rev.*, 106 (1957) 340; B. L. Ioffe, L. B. Okun, and A. P. Rudik, *J.E.T.P.* (*U.S.S.R.*), 32 (1957) 396.
5. We remark here that if the neutrino is described by a two-component theory (see Section III) then the result of the large angular asymmetry in ^{60}Co decay establishes in a trivial way the non-invariance property of β-decay under the charge conjugation operation. However, this non-invariance property can also be proved under a much wider framework. In this section we take as an example the case of a four-component theory of neutrino to illustrate such a proof.
6. For notations and definitions of γ matrices see, e.g., W. Pauli, *Handbuch der Physik*, Julius Springer Verlag, Berlin, 1933, Vol. 24.
7. For a summary of these experiments see, e.g., *Proceedings of the Seventh Annual Rochester Conference*, Interscience, New York, 1957.
8. R. L. Garwin, L. M. Lederman, and M. Weinrich, *Phys. Rev.*, 105 (1957) 1415; J. I. Friedman and V. L. Telegdi, *Phys. Rev.*, 105 (1957) 1681.
9. G. Culligan, S. G. F. Frank, J. R. Holt, J. C. Kluyver, and T. Massam, *Nature*, 180 (1957) 751.
10. C. A. Coombes, B. Cork, W. Galbraith, G. R. Lambertson, and W. A. Wenzel, *Phys. Rev.*, 108 (1957) 1348.
11. J. Crawford, et. al., *Phys. Rev.*, 108 (1957) 1102; F. Eisler et al., *Phys. Rev.*, 108 (1957) 1353; R. Adair and L. Leipuner, *Phys. Rev.*, (to be published).
12. T. Coffin, R. L. Garwin, L. M. Lederman, S. Penman, and A. M. Sachs, *Phys. Rev.*, 107 (1957) 1108.
13. T. D. Lee and C. N. Yang, *Phys. Rev.*, 109 (1958) 1755.
14. H. Weyl, *Z. Physik*, 56 (1929) 330.
15. Cf. W. Pauli, *Handbuch der Physik*, Julius Springer Verlag, Berlin, 1933, Vol. 24, pp. 226–227.
16. The possible use of a two-component theory for expressing the nonconservation property of parity in neutrino processes was independently proposed and discussed by T. D. Lee and C. N. Yang, *Phys. Rev.*, 105 (1957) 1671; A. Salam, *Nuovo Cimento*, 5 (1957) 299; and L. Landau, *Nucl. Phys.*, 3 (1957) 127.
17. The possible existence of a conservation law for leptons has been discussed before the discovery of nonconservation of parity. Cf. E. Konopinski and H. M. Mahmoud, *Phys. Rev.*, 92 (1953) 1045.

Biography

Tsung Dao Lee was born on November 24, 1926, in Shanghai, China, as the third of six children of Tsing Kong Lee, a business man, and Ming Chang Chang.

He was educated at the Kiangsi Middle School in Kanchow, Kiangsi, from which he graduated in 1943. He did his matriculation at the National Chekiang University in Kweichow province. The Japanese invasion forced him to flee to Kunming, Yunnan; here he attended the National Southwest University where he met Chen Ning Yang, who in 1957 was to share the Nobel Prize with him.

Being a most promising student in physics he was, in 1946, awarded a Chinese Government Scholarship, which took him to the University of Chicago, where he gained his Ph. D. degree in 1950 on his thesis *Hydrogen Content of White Dwarf Stars*. For some months in 1950 he served as research associate at Yerkes Astronomical Observatory, Lake Geneva, Wisconsin.

From 1950 to 1951 Dr. Lee was a research associate and lecturer at the University of California in Berkeley, and then accepted a fellowship at the Institute of Advanced Study at Princeton, N. J. Here he was a member of the Institute's Staff, from 1951 to 1953, and had occasion to work jointly with his friend Dr. Yang.

Lee was then fast becoming a widely known scientist, especially for his work in statistical mechanics and in nuclear and subnuclear physics, having solved some problems of long standing and of great complexity. Dr. J. Robert Oppenheimer praised him as one of the most brilliant theoretical physicists then known, whose work was characterized by a « remarkable freshness, versatility, and style ».

Apart from the subject of parity non-conservation – which earned him the Nobel Prize – and statistical mechanics and nuclear physics mentioned earlier, his investigations also comprised field theory, astrophysics, and turbulence.

Lee was in 1953 appointed Assistant Professor of Physics at Columbia University, and afterwards successively promoted to Associate Professor

(1955) and Professor (1956). He was then, at the age of 29, the youngest professor in the Faculty; the next year, being recipient of the Nobel Award at barely 31 years of age, he became the second youngest scientist ever to receive this distinction. (The youngest was Sir Lawrence Bragg who shared the Physics Prize with his father in 1915, at the age of twenty-five.)

Together with Dr. Yang, Lee wrote several prominent articles in *The Physical Review*.

Among the honours bestowed on Professor Lee are: the Albert Einstein Commemorative Award in Science of Yeshiva University, New York (1957) and the Science Award of the Newspaper Guild of New York. He has also been elected Fellow of the American Physical Society and the Academia Sinica, and was awarded a D.Sc. degree by Princeton University (1958).

Lee married (Jeannette) Hui Chung Chin a former university student, in 1950.

His favourite pastimes are: playing with his two young boys, James and Stephen; and reading « whodunits » (detective novels).

Physics 1958

PAVEL ALEKSEJEVIČ ČERENKOV

IL'JA MICHAJLOVIČ FRANK

IGOR' EVGEN'EVIČ TAMM

«for the discovery and the interpretation of the Čerenkov effect»

Physics 1958

Presentation Speech by Professor K. Siegbahn, member of the Swedish Academy of Sciences

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

The discovery of the phenomenon now known as the Čerenkov effect, for which the Nobel Prize is today being awarded, is an interesting example of how a relatively simple physical observation, if followed through in the right way, can lead to important findings and open up new paths to research. Here the scientist must be endowed with that unique intuitive experimental disposition which is the true incentive in scientific progress.

Among the students at the Lebedev Institute in Moscow in the early thirties was Pavel Čerenkov. The task assigned to him by his teacher, Professor Vavilov, for his thesis work, was to study what happens when the radiation from a radium source penetrates into and is absorbed in different fluids. The same problem had no doubt concerned many scientists before this young graduate student and, to be sure, many had also observed the weak bluish glow that emanated from the liquid as the radiation penetrated it. Special mention should be made of the important observation of the Frenchman Lucien Mallet. The bluish glow had – as it seemed on good grounds – always been considered a manifestation of the well-known fluorescence phenomenon. This phenomenon has for more than half a century been used, for instance, by radiologists in X-ray fluoroscopes, where the «invisible» X-radiation is allowed to strike a fluorescent screen, which then lights up.

Čerenkov, however, was not convinced that the light phenomenon he had observed was really of the fluorescence nature. Already his first experiments indicated that his suspicions were correct. He found, for instance, that the radiation was essentially independent of the composition of the liquid. This was in disagreement with the fluorescence explanation. By observing radiation even in doubly distilled water, he eliminated the possibility of minute impurities fluorescing in the liquids.

Čerenkov made the new, unknown radiation the subject of a systematic investigation. In his work he found that the radiation was «polarized» along the direction of the incoming radium radiation and that it was the fast

secondary electrons, produced by the latter, that were the primary cause of the visible radiation. This was verified by irradiating the liquids with only the electrons from a radium source.

The investigations that Čerenkov published in the Russian periodicals between 1934 and 1937 essentially established the general properties of the newly discovered radiation. However, a mathematical description of the effect was still lacking. Here two of Čerenkov's colleagues in Moscow enter into the picture. How can a fast electron on passing through a liquid give rise to radiation with the properties observed by Čerenkov? In the beginning, the phenomenon seemed difficult to understand but in the work of Frank and Tamm (1937) an explanation was given that besides being both simple and clear, also satisfied the requirements for mathematical stringency.

The phenomenon can be compared to the bow wave of a vessel that moves through the water with a velocity exceeding that of the waves. This is, incidentally, a simple experiment that anybody can make. First one drops an object into a bowl of water and observes the propagation velocity of the circular wave front. Then one moves the object along the water surface very slowly to begin with but gradually increasing the velocity. When the latter exceeds the wave velocity previously observed, a bow wave is formed that extends obliquely backwards in the well-known way.

The wave velocity in the water surface is of course low and therefore it is easy to produce the bow wave in this case. In air, an analogous phenomenon occurs when a jet plane penetrates the so-called sound barrier at about 1,000 km/h, i.e. when the jet velocity exceeds the propagation velocity of the sound waves. This is accompanied by a bang.

The condition that is required to form the corresponding Čerenkov bow wave of ordinary light when a charged particle, e.g. an electron, traverses a medium is, analogously, that the particle moves with a velocity greater than that of light in the medium. At first, one might think this is impossible, for according to Einstein's famous theory of relativity the velocity of light is the highest possible velocity. This is in itself correct, but the velocity referred to in Einstein's theory is the velocity of light in empty space or vacuum. In a medium, e.g. a liquid or a transparent solid, the velocity of light is lower than in vacuum and furthermore varies with the wavelength. This fact is well-known from school experiments on the refraction of light in a prism. In such a medium, it is thus entirely possible for an ultra-fast electron, emitted from a radioactive source, to move with a velocity greater than that of light in the medium. In that case, a Čerenkov bow wave is formed and

the liquid glows with the bright blue magic shine from the hectic race of the electrons with the out-distanced light.

A beautiful sight is seen on looking down into a uranium reactor containing water; a so-called swimming-pool reactor. The whole core is aglow with the blue Čerenkov light and in this light one can even photograph the inside of the reactor.

In the successful studies of new elementary particles undertaken during the last few years, e.g. the discovery in 1955 of the antiproton – the negative hydrogen nucleus – the Čerenkov effect has played a decisive part. An instrument based on this effect has been designed that is capable of registering the passage of single particles. Only under the condition that the particle has a sufficiently high velocity will it be registered by the instrument which, at the same time, can measure the velocity. For the velocity determination, which can be made with considerable accuracy, one makes use of the fact that the angle of the bow wave depends on the particle velocity. The faster the particle moves, the smaller will be the angle between them. This is easily understood from the example with the vessel in the water. This new type of radiation detector has been named after Čerenkov and is now among the most important instruments at the big atomic laboratories, where elementary particles are accelerated to extremely high velocities.

The discovery of Čerenkov, Frank, and Tamm, about twenty years ago has thus, during the last few years, found an application of decisive importance in the study of the basic structure and nature of matter.

Professor Čerenkov, Professor Frank, Academician Tamm. The Swedish Royal Academy of Sciences has awarded to you the Nobel Prize for Physics for your discovery and explanation of the effect which now bears the name of one of you. This discovery not only throws light upon a hitherto-unknown physical phenomenon, but also provides a new and effective tool for the study of the atom. I congratulate you most heartily on behalf of the Academy, and ask you to accept the prize from the hands of His Majesty the King.

PAVEL A. ČERENKOV

Radiation of particles moving at a velocity exceeding that of light, and some of the possibilities for their use in experimental physics

Nobel Lecture, December 11, 1958

Research into, and the experimental proof of, the remarkable properties of the radiation arising from the movement in a substance of fast electrically charged particles stretches over a period of nearly twenty-five years. As early as 1934 two reports were published – one by S. I. Vavilov, the other by me[1,2] – in which it was indicated that the gamma rays of radium besides causing luminiscence of solutions, also calls forth a visible, faint light from solvents.

In these reports the universal character of this light and its unusual properties were described and the conviction expressed that the newly discovered light could not, in view of its properties, be a luminescence phenomenon at all.

By further experiments it was established that this light was not released directly by the gamma rays, but through high-speed Compton electrons which appear under the influence of the gamma rays by reason of the Compton effect. Attempts to produce a light with the same properties by the action of X-rays ($h\nu_{max} = 30$ keV) were without result.

To begin with, one might suppose this sort of light from the solvent to be of no especial interest, since luminosity is a widespread phenomenon with very different causes in solids as well as fluids. Apart from the generally known «classical» luminescence phenomena one could, for example, point to the faint light from actually very «pure» fluids which arises by the effect of ultraviolet rays[3]. Many fluids emit light when struck by X-rays[4]. It has even been established that fluids emit light by the action of ultrasonic rays[5]. From the time of Pierre and Marie Curie numerous cases of light emission from fluids and solids by the action of radioactive radiations have been known[6].

As a rule, light emissions of this kind are nothing else than the usual luminescence and, in the case of the so-called «pure» fluids, are produced by the presence of tiny quantities of admixtures capable of luminescence.

Hence we were inclined to think that this light produced by the gamma rays was one of the many luminescence phenomena. Pierre and Marie Curie thought so and they were incontestably among the first to observe this kind of light, at any rate under conditions where it was fairly heavily masked by the ordinary luminescence.

Other observers later took the same view, among them Mallet[7], who had not only observed this light phenomenon but had even photographed its spectrum.

However, a more detailed quantitative investigation of this light process enabled us to find a range of properties so remarkable as to afford incontrovertible proof that here we were dealing with no ordinary everyday luminescence, but with a phenomenon of an entirely new kind; and one of extraordinary interest not only on account of its significance in principle but also in regard to the many practical possibilities for its use.

It would nevertheless be a mistake to believe that such a characteristic phenomenon had not been discovered earlier merely through being accidentally « overlooked ».

The unusual nature of this newly discovered phenomenon could only be investigated by way of quantitative determination of the most important radiation characteristics, and by finding out their dependence on definite experimental conditions.

Nowadays, when researchers have at their disposal powerful sources of fast electrically charged particles and very sensitive measuring devices, measurements of this kind do not involve any especial difficulties. But a few decades ago, the means available to the physicists were nothing like so adequate. In those days absolutely the only sources of electrically charged particles available to them were natural radioactive preparations whose intensity was fairly weak. For this reason the intensity of the light, even in fluids* which had helped to excite it, was so weak that the observer could only perceive the light after he had spent a certain time in complete darkness. It is quite clear that under such conditions there was no question of the usual methods of photometry for quantitative determinations.** To carry out determinations of this kind a new and much more sensitive method was

* Research into the above phenomenon was chiefly undertaken with fluids because fluids are easier to purify from luminescent admixtures. With fluids it is easier to change such parameters as viscosity, refractive index, density, etc. Moreover, in this case the experiments in quenching the luminescence become much simpler.

** We must remember that at that time photomultipliers had not yet been developed.

necessary. In the Physical Institute of the Academy of Sciences of the U.S.S.R. where this phenomenon was discovered we made use of the method of visual photometry depending on the threshold vision developed shortly before by E. M. Brumberg and S. I. Vavilov[8], otherwise called the quenching method. This method makes use of the human eye instead of a light-measuring device*. Since the sensitivity of an eye adapted to darkness is some tens of thousands of times greater than its sensitivity by daylight, this method was superior to others by virtue of its higher sensitivity. Notwithstanding its subjectivity and the comparatively large errors in the measurements, this method was at the time the only one that could be used which permitted a quantitative determination of those extremely low light intensities.

It is of fundamental importance to notice that it was only the transition to quantitative determinations which permitted the unusual attributes of the rays in question to be demonstrated and hence their particular origin to be shown.

It has already been indicated that the first hypothesis, and the most credible of all regarding this light, attributed it to a luminescence phenomenon. Now the correctness of this assumption could only be confirmed if it could be proved experimentally that only those characteristics proper to luminescence were present in this light.

There is indeed a large number of luminescence phenomena which are distinguishable from one another by the excitement method, by the fluorescent lifetime, by the character of the spectrum, by the properties of the luminescent substances, and by other peculiarities. In the case in which we are interested it is not a question simply of establishing the presence or absence of luminescence characteristics; but it is important to determine features which characterize a luminescence phenomenon as such unequivocally and universally.

One of these universal characteristics of luminescence is, as S. I. Vavilov has indicated, the total duration of the excited state ($\tau > 10^{-10}$ sec). This property of luminescence permits of influencing the dying-out process. We can, for example, reduce its brightness considerably or, in other words,

* The quenching method, which is now only of historical interest, depends on the following properties of the human eye:
(a) The fact of a visual threshold.
(b) The permanence of the energy threshold relative to one and the same observer (when the conditions are constant).

« quench » the luminescence either by heating the luminescent solution or by adding materials which are able to quench the luminescence. In both cases there follows a reduction of luminescence as a result of the excited particles giving up energy to those not excited, and by the consequent transformation of energy into thermal energy.

Likewise, the polarization of the luminescence can be altered as one influences the mobility of the particles, e.g. by heating.

Suitable experiments have, however, shown that the intensity of the light of fluids cannot be influenced either by heating or by dissolving in them such active fluorescence quenchers as potassium iodide, silver nitrate, and many others. It has also been shown that the perceptible polarization which appears in this light cannot be altered either. It is of fundamental significance that experiments aimed at quenching the fluorescence of undoubted fluorescent solutions (as, for example, the aqueous solution of aesculin), carried out parallel and under the same conditions, showed in all cases a visible quenching effect.

These results prove a practically inertionless character of the dying-out process and excluded the hypothesis of a luminescence phenomenon. Confirmation was forthcoming also in the unusual character of the polarization of this light. The main direction of the vector of the electric vibrations did not run perpendicularly to the exciting beam, as is the case with polarized fluorescence, but parallel to it.

Taken all together, the results collected even during the first stage consequently gave rise to the statement that the light produced in fluids by the action of gamma rays is not a trivial phenomenon. But these facts were not sufficient for us to build an incontestable theory on this basis. This problem was solved somewhat later after the discovery in 1936 of a new especially remarkable property of this radiation, namely its pronounced asymmetry.

It becomes apparent that the light exhibits a grossly marked spatial asymmetry. This radiation is only emitted forwards in one direction which forms a certain angle with the exciting gamma-ray beam.

The investigation of this fundamental property of the radiation proved to be a decisive step towards clarifying the true physical nature of this phenomenon and propounding a theory. We owe the formation of this theory to I. M. Frank and I. E. Tamm[11].

This theory proceeds from the fact that the light described is produced by the electrons which move uniformly in the substance at a speed exceeding the phase velocity of light in this medium.

It is an interesting fact that as early as 1901 Lord Kelvin[12] maintained that the emission of particles was possible at a speed greater than that of light.

Somewhat later, in 1904 to 1905, shortly before the theory of relativity came into being, Sommerfeld[13] submitted the hypothetical case of the movement of an electron at a speed greater than that of light in a vacuum to a theoretical study.

But the coming of the theory of relativity which affirms that material bodies are unable to move at the speed of light, still less to exceed it, overshadowed Sommerfeld's conclusions which seemed less to the purpose.

It is, seemingly, to this circumstance that we may to some extent ascribe the complete neglect of the problem of the movement of electrically charged particles in a substance, because it could not be reconciled with the theory of relativity.

In the case of the movement of the charge in a substance, velocities exceeding that of light are possible without in any way contradicting the theory of relativity.

This fact is explained thus: The speed of propagation of the light waves in a substance differs from the speed of light in a vacuum by the power n, where n is the refractive index of the medium in which the movements takes place.

As for visible light $n > 1$ and as the speed of the propagation of the light waves in the medium is equal to c/n, this will consequently be smaller than the velocity of light in the vacuum c.

On the other hand it has been known for a long time that the velocity of the beta particles emitted from radioactive substances can come very near to the velocity of light c. In their movement in a substance these particles can possess a speed which is greater than that of light in this substance (c/n) while yet remaining less than c, in complete agreement with the requirements of the theory of relativity. Hence the movement of particles at a speed greater than that of light is not only possible in principle but can also be obtained experimentally.

While we accept that the velocity of an electron moving in the medium exceeds that of light, we can also, on the basis of simple qualitative experiments, determine the prerequisites for the formation of this type of radiation and find out a few of its most important properties. Let us assume that an electron achieves an equal movement in the medium in the direction of the axis z at a speed $v > c/n$. At every point which the electron touches an electromagnetic excitement is set up which is propagated in the form of a delayed wave from these points.

If we consider the components of a certain frequency ω of the waves issuing from the various points at an angle Θ to the path of the electron

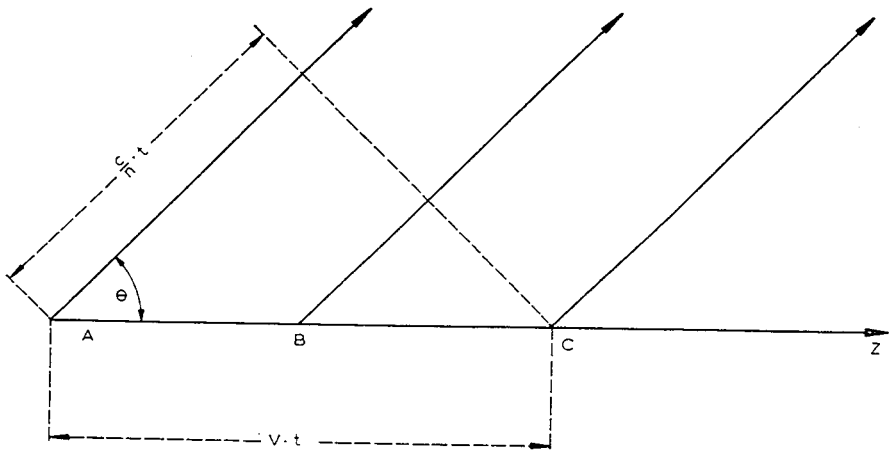


Fig. 1. Mechanism of radiation.

(see Fig. 1), it is easy to satisfy ourselves that in all directions except that in which

$$vt \cos \Theta = \frac{c}{n} t \quad \text{or} \quad \cos \Theta = \frac{1}{\beta n} \tag{1}$$

the waves are quenched through interference. On the contrary, in the direction which satisfies condition (1) the waves reach the observer with an optical path-length difference equal to zero, and accordingly the radiation will only result in this direction.

This radiation has an analogy in acoustics in the form of the so-called shock waves produced by a projectile or an aeroplane travelling at an ultrasonic velocity (Mach waves). A surface analogy is the generally known bow wave.

From Eq. (1), which represents one of the most visible results of the theory of Tamm and Frank, it follows that it is possible for the radiation to be produced only under the condition $\beta n > 1$, i.e. when the velocity of the particle v exceeds the velocity of light c/n.

Consequently, the equation $\beta n = 1$ expresses the energetic threshold of the radiation. The quantity E_0 of this threshold is determined by the refraction coefficient n. Since it is not directly the energy of the particle that is

required for the determination of this threshold, but its velocity, it is therefore obvious that E_o is also dependent on the mass of the particle.

To illustrate the above, the following Table 1 shows the values of the threshold energy E_o for electrons, π-mesons and protons, given three different values for n.

Table 1.

Type of particles	*Threshold energy in MeV*		
	$n = 1.3$ $\beta = 0.769$	$n = 1.5$ $\beta = 0.67$	$n = 2.0$ $\beta = 0.50$
Electrons	0.29	0.2	0.078
π-mesons	79.0	47.0	21.5
Protons	520	320	143

The theoretical interdependence of the quantities Θ, β, and n which is expressed by the relationship (1) was tested experimentally. The results obtained are in complete agreement with the requirements of the theory. Fig. 2 is a diagram of the experiment to determine this interdependence.

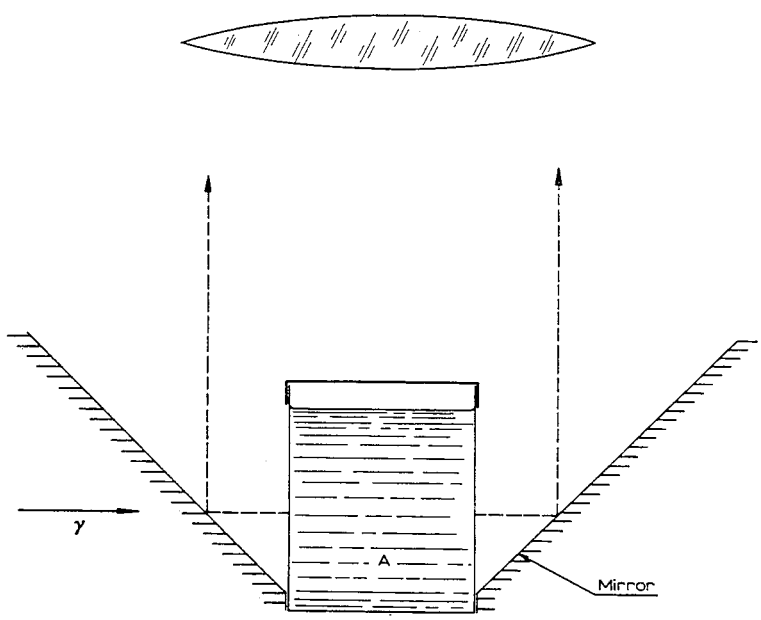


Fig. 2. Diagram of experiment to determine the distribution of intensity according to angle.

The gamma-ray beam falls on a thin-walled vessel filled with fluid (A in Fig. 2). When the radiation arising in this fluid leaves the vessel it falls on a conical mirror and is reflected by this on the objective of a photographic apparatus. The lumisnescence – since it exhibits no asymmetrical properties – produces a picture in the form of a closed ring.

On the other hand, the radiation of particles with a velocity exceeding that of light does not give a closed ring in the image but two patches with an angle between them equal to $2\,\Theta$.

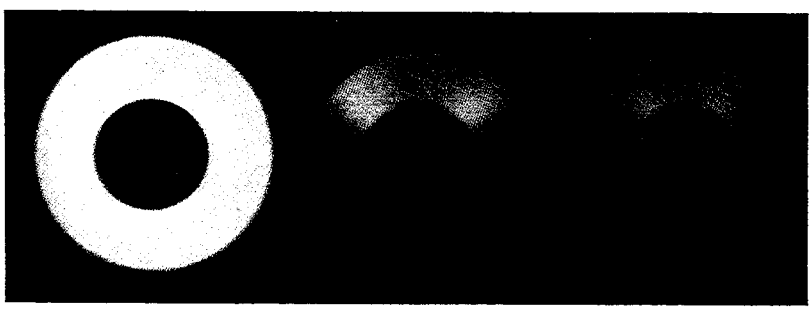


Fig. 3. Photograph of the distribution of intensity according to angle: (*a*) ordinary luminescence (solution of aesculin in water); (*b*) light from ethyl cinnamate ($n = 1.5804$); (*c*) light from water ($n = 1.3371$).

Fig. 3 (*c*, *b*) shows examples of such photographs for two pure fluids (water and ethyl cinnamate). For comparison, a photograph is also shown of the luminescence of an aqueous solution of aesculin (Fig. 3*a*).

The distribution of angles of the radiation intensity (for four fluids) established by measurements from these photographs are set out in Fig. 4. For each of these fluids two curves were plotted corresponding to the excitation of the light by the gamma rays of ThC″ (the upper curves) and by those of Ra (the lower curves).

From Fig. 4 it is easy to determine graphically the angle Θ. The values of this angle increase with the rise in the refractive index n, exactly as it should do in accordance with the theory. Results were obtained for one and the same fluid which in experiments with gamma rays of ThC″ gave higher values of Θ than in experiments using gamma rays of Ra. This difference in measurements in the case of $\Theta_{ThC''}$ and Θ_{Ra} allows us to use Eq. (1) to determine the « effective » velocity ($\beta_{eff.}$) of the Compton electrons which in-

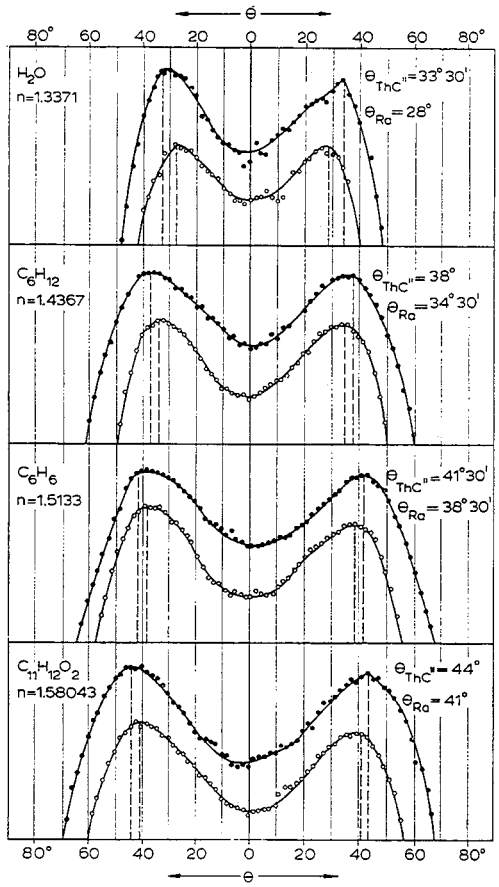


Fig. 4. Distribution of intensity according to angles for fluids with various values for n, obtained from the photograph in Fig. 3. The curves with dark points (*upper curves*) correspond to the excitation of the light by Compton electrons from the gamma rays of ThC″. For these, $\beta_{eff.} = 0.869$ (cf. Eq. (1)). The *lower curves* correspond to the excitation by Compton electrons of the gamma rays of Ra. In this case $\beta_{eff.} = 0.847$.

deed excite the radiation. These velocities amounted to 0.869 and 0.847 respectively. This result is completely in accord with the higher energy of the gamma rays of ThC″.

If we consider the picture not in the plane but spatially, then the radiation must spread out along the surface of a cone whose axis forms the path of an electrically charged particle while the surface line forms with this axis the angle Θ.

If we place the photographic plate perpendicularly to the beam of the high-speed particles (Fig. 5) we shall obtain, in addition to an image of the track of the beam, also a photograph of the radiation in the form of a ring. This photograph (Fig. 6) was obtained with the aid of a fine beam of protons in the accelerator of the United Institute for Nuclear Research at Dubna.

Up till now, we have in our considerations assumed some fixed frequency ω. In reality, however, the radiation spectrum is continuous. Since the medium exhibits dispersion, i.e. the refractive index is dependent on the frequency, this means that the light of different wavelengths is propagated at angles which, even with strictly constant velocity of the particles, differ somewhat from one another.

Thus the radiation is broken up as in spectral analysis. The radiation cone will consequently show a definite intensity, and in the case of a medium with normal dispersion the spectral red will lie in the inner part of the cone while the violet is on the outside.

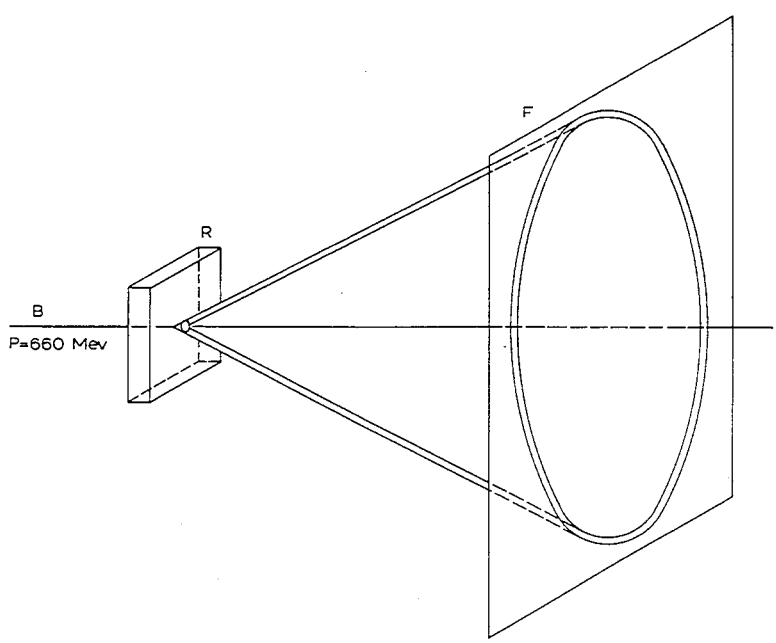


Fig. 5. Diagram of experiment to obtain photographs of the cone section in the plane of the photographic plate.

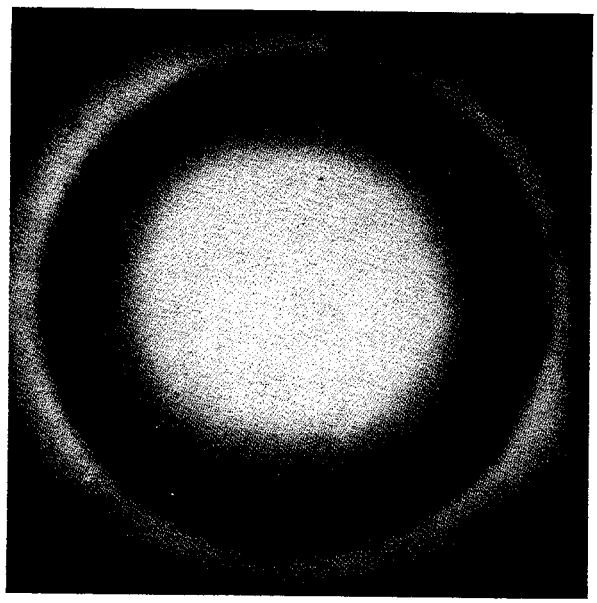


Fig. 6. Photograph of a section of the radiation cone obtained in an experiment of which Fig. 5 is a diagram. The central patch is the track of the proton ray.

[That this is actually so was shown by a photograph showing part of the ring in Fig. 6 taken with a colour plate*.]

Thus the interpretation of the ray mechanism suggested by Tamm and Frank makes clear, even under qualitative consideration, the especially characteristic properties of this radiation, such as, for example, the asymmetry, the short dying-out period, the presence of an energetic threshold and the universal character of this radiation. Moreover the strict quantitative theory supplies an expression for the energy W which the electron gives up during this radiation. The formula is expressed as follows:

$$W = \frac{e^2 l}{c^2} \int_{\beta n > 1} \omega \left(1 - \frac{1}{\beta^2 n_2}\right) \mathrm{d}\,\omega \qquad (2)$$

where l is the length of the path of the electron.

* Unfortunately the coloured illustration could not be reproduced here. The reader is referred to *Les Prix Nobel en 1958*, opposite page 90.

From this it follows likewise that the energy of the radiation spectrum is proportional $1/\lambda^3$, i.e. that it increases powerfully in the direction of the short wavelengths.

In the Röntgen field, however, the radiation must stop because in this field $n < 1$.

Finally it follows from the theory that the light must exhibit a polarization, the same in fact as established in the first experiments; the vector of the electric vibrations lies in the plane which contains the ray and the direction of movement of the particles.

This all goes to show that the theory in question covers exhaustively all the properties of the new radiation known up to now. The development of this theory completes a great cycle of research work which covers discovery, all-round experimental investigation, and the development of the theoretical basis of this phenomenon which established a new branch of physics – the optics of rays moving faster than light.

Interest in the new radiation was indeed great, but, in the absence of sufficiently sensitive and convenient measuring devices, only as a matter of principle. Its potentialities for practical use, especially in experimental physics, remained unexploited.

In recent years, however, following the development and production of photomultipliers, the radiation from high-speed electrically charged particles acquired important practical significance especially in the field of research into the physics of high-energy particles.

Although the intensity of the flash called forth by a single particle is extremely low it can now be measured. Formula (2) shows that when $\beta \simeq 1$, the number of photons which an electrically charged particle moving in a medium where $n \simeq 1.5$ emits, and which fall within the visible part of the spectrum, will amount to 200–300 photons per cm. If we choose the right form and position of the radiator (i.e. the medium in which the electrically charged particle occasioning the radiation moves) an important part of this light can be conducted to the cathode of the photomultiplier. As a result of a multiple intensification at the anode of the multiplier, a current impulse arises which is millions of times stronger than the current we started with. By means of a suitable radiotechnical scheme, this impulse can be maintained and the particle thus observed. A device for this would be a counter which would measure the particles by means of the radiation proceeding directly from them.

This kind of counter is highly reminiscent of the so-called scintillation counter. In this, the measurement of electrically charged particles, which is likewise done with the aid of a photomultiplier, makes use of the luminescence arising from the absorption of energy from the particles in the scintillator. But compared with the scintillation counter, this one has many important advantages:

1. The short duration of dying-away permits the development of counters with a very high resolving power.
2. The presence of an energy threshold makes this type of counter insensitive to the slow particles whose energy lies below the threshold. This property of the counters is especially valuable in those cases where there is a considerable gamma radiation level.
3. The asymmetry of the radiation facilitates with this counter the measurement of those particles which are moving in the radiator in the direction of the cathode of the photomultiplier. Particles moving in the opposite direction are not measured by the counter. In other words, this type of counter has the distinction of focussing on a definite directional effect.

Winckler[14] made use of this characteristic of the counter in determining the «albedo» of the cosmic rays in the upper layers of the atmosphere.

At the present time a large number of counters of this sort is described in the works of Jelley, Marshall, and other authors[15], and are distinguished by their original construction.

The practical value of the radiation of high-speed particles does not lie solely in the possibility of using it to detect particles. The full utilization of the peculiar properties of this radiation (often in combination with other methods) considerably enlarges the possibilities of physical experiment in a wide range of cases.

Thus, for example, it is known that one of the most important parameters of a particle – its mass – can be determined on the basis of measurements of its momentum and speed. In the measurement of speeds one usually comes up against experimental difficulties. It is evident without more ado that where the speed of a particle is within a certain range where β (that fulfills the condition $\beta n > 1$) is still sufficiently different from 1, the speed of the particle, on the basis of Eq. (1) and starting from the measured quantity Θ and the known refractive index, can easily be calculated.

When the type of the particle is known, speed measurements enable us

to determine their energy also. This method produces particularly good results in the case of energies in the region of several hundred MeV (see Table 1). These energy measurements are accurate to 0.25%[16].

We have already noted the fact of the radiation having an energetic threshold renders the counter insensitive to low-energy particles. We thus have the possibility of altering the threshold energy E_o when we choose a radiator with a suitable quantity n.

It is clear that two counters set at degrees of threshold energy E_o' and E'', and switched on in the proper sequence according to the scheme of anti-coincidence will only measure those particles whose velocity is in the region of E_o' and E_o''.

A procedure of this kind was successfully employed by Serge and his collaborators in their outstanding work which led to the discovery of the antiproton.

Another interesting field of application of the properties of the rays had been found in their use for investigating both showers of cosmic rays. In the investigation of these showers with the aid of devices with ordinary counters, at a certain altitude only those particles are measured which are remotely « descended » from the primary particle. The devices did not measure all the other particles which had emerged during the previous developmental stages of the shower and had already disappeared. But since the shower particles possess gigantic energies they have the power of producing in the air a radiation of the type in question which in practice propagates itself in the same direction. Under favourable conditions this radiation reaches the surface of the earth and can be measured by a photomultiplier. This method procures a more complete picture of the shower and permits valuable knowledge of its developmental process to be obtained.

The distribution in the primary cosmic radiation (outside the atmosphere) of nuclei heavier than hydrogen nuclei is of great importance in cosmology. Investigations into this are at present being carried out via the sputniks. These enquiries are based on the fact that the intensity of the radiation of particles travelling at a velocity greater than that of light is proportional to the square of the charge of the particle. Hence, the impulses coming from particles variously charged, and registered by the counter, differ according to the amplitudes. Analysis of the distribution of amplitudes will allow us to obtain information about the distribution of heavy particles in the cosmic radiation according to their ordinals.

The last subject I would like to mention is the use of the radiation of high-

speed particles for the purpose of measuring their energy, if this energy is fairly high. Here, the measurement of particle energy on the basis of deflection in the magnetic field is no longer possible. But we can attempt to determine it by measuring the total energy which the particle gives up to the radiation of the type mentioned. For this purpose highly transparent, dense radiators must be used which are sufficiently strong, and allow a complete development of the showers.

In this case water proves to be a suitable radiator. In the Institute of Physics in the Academy of Sciences of the U.S.S.R. a device is being constructed for measuring the energy of cosmic particles by this method.

The examples set forth show the great importance which the radiation caused by particles moving at a speed greater than that of light has acquired in experimental physics.

Even so, we have not by a long way exhausted all the possibilities for their practical use. There can be no doubt that the usefulness of this radiation will in the future be rapidly extended.

1. P. A. Čerenkov, *Dokl. Akad. Nauk SSSR*, 2 (1934) 451.
2. S. I. Vavilov, *Dokl. Akad. Nauk SSSR*, 2 (1934) 457.
3. S. I. Vavilov and L. A. Tumermann, *Z. Physik*, 54 (1929) 270.
4. Z. S. Newcomer, *J. Am. Chem. Soc.*, 42 (1920) 1997.
5. Z. Frenzel and Z. Schultes, *Z. Physik. Chem.*, B 27 (1934) 421.
6. Eve Curie, *Madame Curie*, London, 1941.
7. M. L. Mallet, *Compt. Rend.*, 183 (1926) 274; 187 (1928) 222; 188 (1929) 445.
8. E. M. Brumberg and S. I. Vavilov, *Izvest. Akad. Nauk. Omen*, Ser. 7 (1933) 919.
9. P. A. Čerenkov, *Dokl. Akad. Nauk SSSR*, 3 (1936) 413.
10. P. A. Čerenkov, *Dokl. Akad. Nauk SSSR*, 14 (1937) 99.
11. I. E. Tamm and I. M. Frank, *Dokl. Akad. Nauk SSSR*, 14 (1937) 107.
12. Lord Kelvin, Nineteenth Century, Clouds over Dynamical Theory of Heat and Light, *Phil. Mag.*, July (1901).
13. A. Sommerfeld, *Göttingen Nachr.*, 99 (1904) 363; (1905) 201.
14. J. Winckler, *Phys. Rev.*, 85 (1952) 1054.
15. J. V. Jelley, *Čerenkov Radiation and its Applications*, Pergamon Press, 1958.
16. R. L. Mather, *Phys. Rev.*, 84 (1951) 181.

Biography

Pavel Aleksejevič Čerenkov was born in Voronezh Region on July 28, 1904. His parents, Aleksei and Mariya Čerenkov, were peasants. He graduated from the Physico-Mathematical Faculty of Voronezh State University in 1928, and in 1930 he took a post as senior scientific officer in the P. N. Lebedev Institute of Physics in the U.S.S.R. Academy of Sciences. He was promoted to section leader, and in 1940 he was awarded the degree of Doctor in Physico-Mathematical Sciences. In 1953 he was confirmed in the academic rank of Professor of Experimental Physics, and since 1959 he has controlled the photo-meson processes laboratory. He has taught in institutes for higher learning for fourteen years.

It was in 1934, whilst he was working under S. I. Vavilov, that Čerenkov observed the emission of blue light from a bottle of water subjected to radioactive bombardment. This «Čerenkov effect», associated with charged atomic particles moving at velocities higher than the speed of light, proved to be of great importance in subsequent experimental work in nuclear physics and for the study of cosmic rays. The Čerenkov detector has become a standard piece of equipment in atomic research for observing the existence and velocity of high-speed particles, and the device was installed in Sputnik III. He has shared in the work of development and construction of electron accelerators and in investigations of photo-nuclear and photo-meson reactions.

Čerenkov was awarded State Prizes in 1946 (with Vavilov, Frank, and Tamm) and in 1951.

In 1930 he married Marya Putintseva, daughter of A. M. Putintsev, Professor of Russian Literature. They have a son, Aleksei, and a daughter, Elena.

Il'ja M. Frank

Optics of light sources moving in refractive media

Nobel Lecture, December 11, 1958

Peculiarities of radiation in a medium

For a number of years the Vavilov-Čerenkov effect appeared as but a peculiar optical phenomenon difficult to observe. Light emission was induced by using radioactive preparations and the glow was observed visually[1]. The weakness of the glow seemed to preclude any application of the phenomenon in physics, and so much the more in engineering.

Since the theory of the Vavilov-Čerenkov effect appeared[2], the phenomenon could be regarded as an instance of super-light velocity optics*. This was a singular example in this field, which seemingly was isolated from any other known physical phenomenon. It was evident that in principle other manifestations of super-light velocity optics were also possible, but their observation appeared very complicated. For example, the first calculations already indicated that if the Vavilov-Čerenkov radiation were induced not by an electric charge, but, say, by the magnetic moment of an electron, it should be so weak that its experimental detection would not be feasible[3]. It was likewise evident that it would be difficult to create conditions for observation of atoms moving at super-light velocities[4].

Theoretical analysis of all these problems was for a number of years of interest chiefly from the viewpoint of principle.

Progress in nuclear physics and the improvement of experimental techniques in recent years has resulted in the fact that the Vavilov-Čerenkov effect has found numerous applications in the physics of high-energy particles. A connection between this phenomenon and many other problems has also been found, as, for example, the physics of plasma, astrophysics, the problem of radio wave generation, the problem of acceleration of particles, etc.

A broader approach to the treatment of the phenomena related to the

* A summary of the results of theoretical work and list of references are given in the review by B. M. Bolotovsky[8].

Vavilov-Čerenkov effect has now become not only justified but essentially necessary.

The question naturally arises as to the peculiarities of a radiation which may be set up not only by an electric charge, but by any source of light, moving in a refractive medium[5]. Such a general approach to the problem, involving, notably, the Vavilov-Čerenkov effect, is of interest now not only from the viewpoint of principle. It may be hoped that some phenomena of this range will also become in the near future a subject of experimental study.

Since the discovery of the Vavilov-Čerenkov effect, our ideas of the mechanism of interaction between a rapidly moving particle and a medium have undergone a considerable change.

Formerly it appeared unquestionable that radiation arising during an electromagnetic interaction between high-energy particles and a medium is always some kind of a « bremsstrahlung ». Most of the energy of such radiation is carried by high-energy photons. The optical properties of the medium should not be of significance for the emission and propagation of such photons. It was also assumed that the processes of ionization and excitation by fast particles might be regarded as a sum of independent interactions of such particles with individual atoms and molecules. This led to the deduction that generally for interaction between high-energy particles and a substance its macroscopic properties are likewise of no importance.

The discovery and interpretation of the Vavilov-Čerenkov effect, and then the connection between this phenomenon and ionization losses, found by Fermi[6], have led to a revision of this viewpoint. It has now become evident that the macroscopic properties of the medium play an important part in the processes of radiation of light by rapidly moving particles.

The ratio between the velocity of the emitter and that of light is a highly important factor on which radiation depends. In a vacuum, the velocity of light is constant and always exceeds that of the emitter. It enters the formulae determining the radiation, as a universal constant. Radiation in a vacuum is therefore determined solely by the nature of the emitter and the law of its motion. The case is different in a refractive medium. The phase and group velocities of light differ from those in a vacuum. They depend on the properties of the medium and on the frequency of the light. In optically anisotropic media, they are a function of the direction of propagation and polarization of the waves. In media of limited dimensions, changes in the velocity of light during transition through the boundary of the media are also of

importance. Hence, in a refractive medium, the ratio between the velocity of the emitter and that of wave propagation depends considerably on the velocity of light in a medium and on its changes. Unlike a vacuum, the ratio may, notably, exceed unity. As a result, not only radiation properties but sometimes even the very fact of its origination depend on the peculiarities of light propagation in a medium. The Vavilov-Čerenkov effect is a case in point.

Radiation in a medium naturally also depends to a very great extent on the nature of the emitter. The theory makes it possible to foretell the properties of the Vavilov-Čerenkov radiation not only for a moving electric charge, but also for other cases. For instance, similar to an electric charge, the Vavilov-Čerenkov radiation should have also been produced by a magnetic charge, had it been proved to exist[7].

Whereas the question of radiation of a magnetic charge should now, too, be considered as being only theoretically possible the question of the Vavilov-Čerenkov effect for magnetic and electric dipoles and multipoles is quite real at present.

As a matter of fact, analysis of the radiation of a moving system of particles may prove necessary in resolving the numerous tasks related to processes in plasma and to problems of acceleration of particles. It is evident that a system of particles may, notably, be quasi-neutral, but it may possess an electric and, particularly, a magnetic moment due to moving ring currents.

A system of particles may not only move as a whole; it may also have natural frequencies of oscillations. This is true to an even greater extent of such systems as a moving atom, ion or atomic nucleus. An electron moving in a magnetic field may likewise possess natural frequency (Larmor frequency of revolution about the lines of a field). Therefore, apart from generalizing the theory of the Vavilov-Čerenkov effect, analysis is also required of the general case of radiation of systems possessing natural frequencies of oscillations[5].

Such a general analysis also includes the Vavilov-Čerenkov effect. The latter corresponds to the limiting case when the natural frequency is zero.

The fact that the theory of radiation of a charge moving with a velocity exceeding that of light has not been revised in the past twenty years does not mean at all the theory of this effect has been fully consummated. This can be seen from the following example. L. I. Mandelstam was the first to point out that it is not necessary for a charge to move in a continuous medium in order to radiate during super-light velocity*. The radiation re-

* See article by V. L. Ginzburg and I. M. Frank[9].

mains the same if the charge moves along the axis of a hollow cylindrical channel inside the medium, provided the diameter of the channel is small in comparison with the length of the emitted wave. For practical purposes this is very important, since it makes it possible to obtain radiation in a medium under conditions when the emitter does not collide directly with the atoms of the medium, which may deform or destroy it. It seemed that this applied also to the radiation of a dipole in a medium.

As recently shown, however, by V. L. Ginzburg and his associates, this question is not so simple as it appeared before[10]. The properties of a medium directly adjacent to the dipole may play an important part, and the presence of a channel of any, even the smallest, diameter cannot, therefore, be ignored.

This important factor has called for a critical analysis of the formerly obtained data as well. Thus, two contradictory results were obtained by two different methods for the radiation of a magnetic dipole[4,7]. It may now be assumed that this was not due to the erroneousness of one of the methods used, but to the fact that they differently took into account the effect of the medium adjacent to the moving dipole. Possibly both results are correct, but they apply to different physical cases. The matter requires, however, further consideration.

The series of problems dealt with in this paper, despite their diversity, comprises but the simplest case of radiation in a medium, namely radiation during which the translational motion of the system may be regarded as uniform and rectilinear.

Transition radiation

A typical example of radiation in a medium and, notably, during the uniform motion of an electric charge, is provided by the so-called transition radiation. The assertion that there is no radiation during a rectilinear and uniform motion of an electric charge at a velocity smaller than the phase velocity of light is correct only under the condition that the velocity of light along the path of the particle remains unchanged. For example, if a uniformly moving charged particle crosses the boundary of two media with different indices of refraction, there appears transition radiation. Radiation appears because the jump which the magnitude of the phase velocity of light undergoes at the boundary of two media is to some extent equivalent to the

jump in the magnitude of the velocity of a particle. The equivalence to bremsstrahlung becomes complete in an extreme case, when the particle moves from vacuum to a metal in which light is absorbed over a length smaller than the wavelength of the light. The intensity of the transition radiation is at its maximum in this case. In the optical range of the spectrum – the only region in which transition radiation occurs – the spectrum and magnitude of the radiated energy are then exactly identical to those of the radiation which would have been produced by an electric charge and a charge of the opposite sign, moving towards it (its electric image in the metal), and which stop instantaneously at the point of encounter.

The intensity of transition radiation at low velocities is proportional to the kinetic energy of the particle, and it rises in the region of relativistic velocities as the logarithm of the total energy. Like bremsstrahlung, it becomes sharply directed in this case. It has been suggested that transition radiation might be useful in determining the energy of ultra-relativistic particles. This is important because it is very difficult to use for this purpose the Vavilov-Čerenkov effect for ultra-relativistic particles. As is well known, the angle at which the Vavilov-Čerenkov radiation is directed, and its intensity, attain in this case a practically constant value.

The use of transition radiation is, however, impeded by the fact that its intensity is very low. The probability of emission of a photon is of the order of the fine structure constant, i.e. of the order of a hundredth. If it is not possible to sum up transition radiation from many plates, observation of an individual particle by transition radiation may be carried out with but little efficiency. In this connection we should like to note the peculiarities of transition radiation at ultra-relativistic velocities. Unlike particles with a low velocity, transition radiation is almost the same during the incidence of such a particle from vacuum on a transparent dielectric as during the incidence on a metal. This is easy to understand by analogy with bremsstrahlung. Indeed, a change in the velocity of light is equivalent to a slight change in the velocity of the particle. But even a small change in the velocity of an ultra-relativistic particle means a great change in its energy, i.e. great deceleration of the particle. This peculiarity may permit the summing of transition radiation from the surfaces of many parallel transparent plates in a vacuum.

The second peculiarity consists in the fact that at ultra-relativistic velocities, the equilibrium field entrained by the particle in a vacuum is formed along a considerable path length. Consequently, to prevent the intensity of radiation from being reduced, the vacuum layers between the plates should

not be less than some preset magnitude. For instance, for a proton with energy of 10^{11} electronvolts, this minimum distance is of the order of 1 mm, which is reasonable; but for a proton with energy of 10^{14} electronvolts it rises to the unreasonable magnitude of a kilometer.

I have dwelt on the subject of transition radiation in order to emphasize the peculiarity of the optical phenomena for radiation sources moving in refractive media, which so greatly depends on the peculiarities of propagation of light in a substance.

It should be noted that although the theory of transition radiation was developed by Ginzburg and the author of this lecture[11] more than ten years ago, and has since been analysed in a number of works* it has not yet been studied experimentally. The situation in this case is almost the same as in the case of the Vavilov-Čerenkov radiation before their papers were published. There is no doubt that transition radiation has also been observed on numerous occasions by various physicists, since the glow of the surfaces of electrodes under the impact of bombarding particles is well known. But even today the part played in this glow by luminescence, bremsstrahlung, and transition radiation has not been elucidated. The most reliable data on transition radiation have recently been obtained by A. E. Chudakov. Using the coincidence method, he observed photons emitted from the surface of a metal foil during the incidence on it of fast electrons from radiophosphorus. The intensity of radiation thus found proved to coincide with the estimated intensity for transition radiation, at least in the order of magnitude**.

It is also worth mentioning that transition radiation is practically always an intrinsic part of the Vavilov-Čerenkov radiation due to the limited thickness of the radiator. As shown by V. E. Pafomov for a radiator of very small thickness this factor should be taken into account[16].

* See, for instance, the papers by Garibyan and Pafomov and the references cited therein[12].

** In the book by Jelley, *Čerenkov Radiation*[18], with which I had the opportunity of becoming acquainted after this paper had been written, there is mention of the fact that in 1958 the author together with Elliott and Goldsmith observed a radiation emitted by 1.5 MeV protons incident on a polished aluminium target. On basis of the data on the intensity and polarization, the investigators concluded that the glow was transition radiation.

Radiation spectrum and quantum interpretation of the phenomenon

The radiation of a charged particle uniformly moving at a velocity exceeding that of light may, as is well known, be fully described by the methods of classical electrodynamics. The quantum theory of this phenomenon was first developed by Ginzburg[3] and then by many other investigators*. Ginzburg has shown that the classical formula for the cosine of the angle at which radiation occurs is correct up to a very small correction of the order of magnitude of the ratio between the energy of the radiated photon and the total energy of the moving emitter. (Even for an electron the ratio is less than 10^{-5}.) If this slight quantum correction contained in the exact formula is disregarded, identical relations between the frequency of the radiated light and the direction of its emission are obtained by both the classical and the quantum methods. Let us write them down in a quantum form, for a system possessing a natural frequency ω_0[14,5] ,where ω_0 is the frequency in the laboratory system of coordinates, that is, $\omega_0 = \omega_0' \sqrt{1 - \beta^2}$.

There is no necessity of assuming in this case that ω_0 is the only natural frequency possessed by the system. It may be regarded as a component of a complex spectrum of frequencies and it should be sufficient to study the radiation related to this frequency.

If the momentum of the photon, which in a medium should be assumed to equal $n\hbar\omega/c$, is very small in comparison with that of the emitter, then the law of momentum conservation during radiation may be expressed as follows:

$$\frac{n\hbar\omega}{c}\cos\Theta = \frac{\Delta E}{v} \tag{1}$$

where ΔE is the change in the kinetic energy of the emitter, and v is its velocity. From their ratio we obtain the magnitude of the change in the momentum of the system.

The change in kinetic energy is apparently determined by the energy of the radiated photon $\hbar\omega$ and the change in the internal energy of system $\hbar\omega_0$

$$\Delta E = \hbar\omega \pm \hbar\omega_0 \tag{2}$$

The term $\hbar\omega_0$ should be taken with a minus sign if, when emitting the pho-

* See, for example, review[8].

ton, the system passes from an upper energy level to a lower one, that is, if the energy of the emitted photon is supplied, partly at least, from excitation energy. The plus sign should be used if the system becomes excited in the process of emission, i.e. if the kinetic energy is spent both on radiation and excitation.

By combining Eqs. (1) and (2), we obtain

$$\frac{n\omega}{c}\cos\Theta = \frac{\omega \pm \omega_o}{v} \tag{3}$$

Factor $\hbar$ has been cancelled out and the equation does not, indeed, contain anything of a specifically quantum nature. The same result is also obtained from classical wave analysis.

In Eq. (3) we can distinguish three cases:

Case 1 – Let us assume that

$$\frac{n\,v}{c}\cos\Theta = 1 \tag{4}$$

Then Eq. (3) is satisfied only if $\omega_o = 0$. This is precisely a case of the Vavilov-Čerenkov radiation, while (4) is a well-known condition determining the direction of emission of light for this radiation. The natural frequency $\omega_o = 0$ required for bringing into effect (4) means that the moving system should contain a source of a time-independent electromagnetic field (an electric charge, a constant dipole moment, etc.). Consequently, for the Vavilov-Čerenkov radiation to take place it is necessary that the constant component of the field should differ from zero. In this case Eq. (4) yields the relation between angle Θ and the radiated frequency, inasmuch as the refraction index $n(\omega)$ is a function of frequency.

Case 2 – Suppose now that the left-hand side of Eq. (4) is less than unity. Then Eq. (3) may be satisfied only if ω_o has a minus sign, i.e.

$$\frac{n\omega}{c}\cos\Theta = \frac{\omega - \omega_o}{v} \qquad \frac{vn}{c}\cos\Theta < 1 \tag{5}$$

This is nothing else but the Doppler condition for a source of light, moving in a medium. It has already been obtained by Lorentz when studying the optics of moving media.

Eq. (5) may evidently be expressed in the following ordinary way

$$\omega = \frac{\omega_0}{1 - \frac{vn}{c}\cos\Theta} \tag{5a}$$

It determines the frequency when the component of the velocity along a ray, $v \cos \Theta$, is less than the phase velocity of light c/n for frequency ω.

Eqs. (5) or (5a) differ from the usual Doppler condition for a source of light moving in a vacuum only in that the velocity of light in a vacuum has been replaced by the phase velocity c/n. If v is small in comparison with the phase velocity of light, and the dispersion of light is not great in the range of frequencies close to ω_0, this does not lead to anything fundamentally new. There is only a change in the absolute magnitude of the Doppler shift. It is so obtained as if it had been in a vacuum for a velocity equal to nv, i.e. n times greater. If the dispersion of light in the medium is great, there arise important peculiarities. The presence of dispersion should not be ignored in any medium when the velocities of motion are comparable to the phase velocity of light. Indeed, with n = constant and for angle $\Theta = 0$, the quantity $(vn/c) \cos \Theta$ would tend to unity with an increase in v while ω, as can be seen from (5a), would tend to infinity. At still greater velocities, the inequality sign in (5) would not be valid and, consequently, (5) would have no solution. As a matter of fact, the refraction index of any media becomes practically equal to unity at sufficiently large values of ω. Hence, the Doppler frequency in this case is the same as it would have been in vacuum, i.e. it is certainly finite. In other words, at any velocity v and any value of Θ, Eq. (5) will have a solution. Moreover, as will be shown below, there may be not one but several solutions[4,5] («complex» Doppler effect).

Case 3 – The third case takes place when the left-hand side of Eq. (4) is greater than unity. Then a plus sign should be before ω_0 in Eq. (3), and thus

$$\frac{n\omega}{c}\cos\Theta = \frac{\omega + \omega_0}{v} \qquad \frac{vn}{c}\cos\Theta > 1 \tag{6}$$

This is a generalization of Doppler's formula for the case[4,14] when the velocity of the emitter exceeds the phase velocity of light for a radiated frequen-

cy*. It determines the «super-light» Doppler frequencies. Like the Vavilov-Čerenkov effect, the super-light Doppler frequencies appear when the velocity exceeds some threshold velocity. They are radiated simultaneously with ordinary frequencies, but only at sufficiently high velocities and within some range of acute angles.

It can be seen from the above quantum analysis that the plus sign at ω_0 in (2) and (6) respectively means excitation of the system. Hence, radiation of super-light photons occurs not during the transition from the upper, i.e. excited state into the lower, as in a general case, but quite the contrary, from the lower into the upper state, the energy being supplied from the kinetic energy of the translational motion of the system[14]. Such a radiation, accompanied by excitation of the system should take place spontaneously if the system is in the lower energy state. This is likewise possible as a spontaneous transition of the system from the upper energy state into the lower, accompanied by emission of photons with a frequency satisfying (5). As a matter of fact, the transition occurs in either case between the same energy states, and the question as to which of them takes place spontaneously is wholly determined by the initial state and the requirements of the conservation laws. In this case Eqs. (5) and (6) are in equal degree consequences of these laws.

The question regarding Doppler's effect in a refractive medium may also be considered within the framework of classical physics. From the viewpoint of classical physics, the results are interpreted as follows. Oscillations with natural frequency ω_0 bring about the appearance of radiation with frequencies which depend on the direction of propagation. It forms a spectrum of Doppler frequencies, which may be of two types. There is always a spectrum of radiation with frequencies satisfying Eq. (5), whose reaction on the emitter causes its damping. Under certain conditions, another spectrum with frequencies meeting Eq. (6) appears in addition to the first. The reaction of radiation of these frequencies promotes the building-up of oscillations. If damping prevails over building-up, oscillations will not arise by themselves in a system for which the classical formulae are correct, and if they existed at the beginning, they will attenuate.

In a quantum system the situation is fundamentally different. The processes of quantum radiation should be considered separately for spectra of both

* Apparently Eq. (6) may be put down in a form similar to (5a). The difference consists only in that the sign in the denominator of the right-hand part of Eq. (5a) should be changed.

types. Therefore, if a process corresponding to Eq. (6) is possible, it is certain to take place, i.e. the system will become excited owing to its own kinetic energy, radiate light, and pass in the usual way to the lower state. In principle, a two-photon mechanism is also possible, photons of both types being radiated simultaneously. Hence, as in the Vavilov-Čerenkov effect, a system possessing a natural frequency of oscillations will spend its kinetic energy on radiation at super-light velocity[14,15].

This can be formulated in the following way: as is well known, motion at a velocity greater than that of light is impossible in a vacuum. It is possible in a medium, but Nature does not lift its ban completely. Any system capable of interacting with radiation will slow down at a super-light velocity by radiating light.

Radiation thresholds

It is evident from the above analysis that the radiation spectrum is determined by the velocity of motion of the system v, its natural frequency ω_0 and the phase velocity of light c/n in a medium in which the radiation is emitted. Both the Vavilov-Čerenkov effect and the Doppler super-light effect are possible, as can be seen from (4) and (6) if $vn(\omega)/c > 1$. This obvious condition for the threshold of their appearance means that the velocity of motion should exceed the phase velocity of light.

This statement, correct for an isotropic medium, determines the threshold of emission of light of a given frequency ω for which the refraction index equals $n(\omega)$. As the refraction index depends on frequency, the threshold is different for another ω. This justifies raising the question in another way: under what condition do the Vavilov-Čerenkov effect and Doppler super-light effect generally become possible in a given medium?*

During radiation in a medium there is yet another peculiarity which likewise appears under certain threshold conditions. It consists in the following.

* For the Vavilov-Čerenkov radiation in an isotropic medium, this point regarding the threshold is elementary, since the latter is determined simply by the maximum value assumed by the refraction index in the given medium. Of importance for further consideration is the fact that for a frequency corresponding to n_{max}, the phase and group velocities are equal (see Eq. (10)), it being evident that for n_{max}, $dn/d\omega = 0$. Hence the fact that the threshold velocity of motion is equal to the phase velocity means that it is also equal to the group velocity of light.

Eq. (3) and, naturally, its sequels (4), (5), and (6), are not linear with respect to ω. As a matter of fact, they contain the refraction index $n(\omega)$ which is a function of the radiated frequency. As a result, not one but several values of ω, satisfying (3) are possible in some cases for given values of Θ, v, and ω_0. This means that several components of different frequency may be radiated simultaneously in a given direction. The appearance of such additional frequencies, i.e. of the so-called complex effects of radiation, is possible only under certain conditions. They may arise not only in the superlight Doppler effect and Vavilov-Čerenkov's radiation, but also in the ordinary Doppler effect subordinated to Eq. (5).

L. I. Mandelstam was the first to draw attention to the fact that the condition under which the complex Doppler effect appeared[4] was related to the magnitude of the group velocity of light. The statement proved to be of a general nature[5].

If we consider radiation in the direction of motion, then in all the enumerated cases the condition for appearance of the radiation or of its new components is that the velocity of the emitter should equal the group velocity of light for a frequency which may radiate (i.e. which satisfies condition (3)). This threshold frequency should evidently satisfy Eqs. (4), (5), or (6), depending on the kind of radiation under consideration.

It is well known that in a refractive medium the transfer of radiation en-

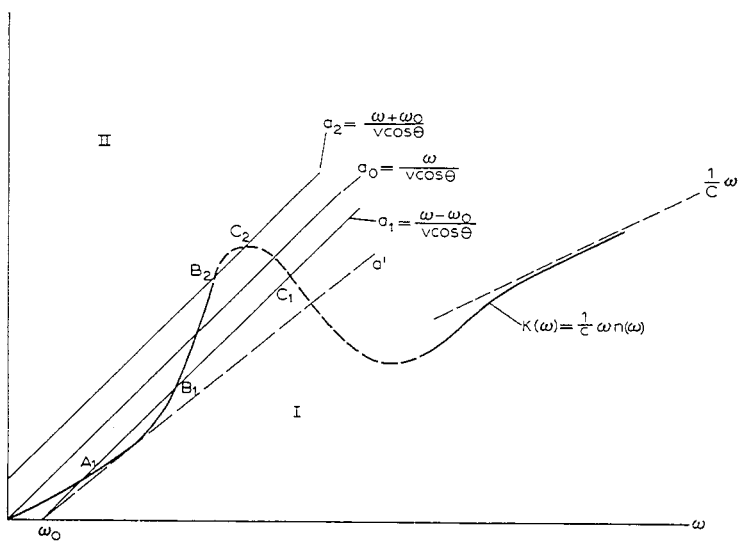


Fig. 1.

ergy occurs not with the phase but precisely with the group velocity. Thus it is not surprising that the group velocity of light is of importance for the processes of radiation in a medium.

The fact that the radiation threshold is connected precisely with the group velocity can be explained by some simple qualitative considerations. Let us assume that the conditions for appearance of the radiation have been fulfilled. Radiation arises and carries energy away from the emitter. Suppose furthermore, that the velocity of motion changes, and approaches the threshold velocity. When the threshold is attained, the radiation should disappear, i.e. removal of energy from the emitter ceases. When the velocity of motion equals the group velocity of light, this will actually take place, since there occurs simply a transfer of energy together with the emitter.

The condition of appearance of the complex effect may be easily determined by analysing the chart in Fig. 1. The curve in Fig. 1 represents dependence of the magnitude of wave vector $k(\omega) = \omega n(\omega)/c$ on the frequency for some imaginable medium. In addition to curve $k(\omega)$, Fig. 1 contains three straight lines whose equations are

$$a_0 = \frac{\omega}{v \cos \Theta} \tag{7}$$

$$a_1 = \frac{\omega - \omega_0}{v \cos \Theta} \tag{8}$$

$$a_2 = \frac{\omega + \omega_0}{v \cos \Theta} \tag{9}$$

The points where the straight lines cross the curve seem to determine at once the frequencies satisfying Eqs. (4), (5), and (6) respectively.

The tangent of the angle of incline of the straight lines a_0, a_1, a_2 to axis ω apparently equals $1/v \cos \Theta$. Let us assume, in accordance with Fig. 1, that $\cos \Theta \geqslant 0$, that is, $\Theta \leqslant \pi/2$.

The nature of intersection of the straight lines a with curve $k(\omega)$ may differ. If we move along the straight line in the direction of increased ω, the straight line may go over at the point of intersection from the region underlying the curve (Region I) into the region above the curve (Region II). This takes place if the slope of the tangent to curve $k(\omega)$, i.e. $dk/d\omega$, is less than $\gamma = 1/v \cos \Theta$ (see, for example, point A_1 on the straight line a_1).

On the contrary, if $dk/d\omega > 1/v \cos \Theta$ then there is a transfer from Region II into Region I at the point of intersection. Finally, $dk/d\omega = 1/v \cos \Theta$ takes place at the point of tangency.

As can be easily proved, the slope of the tangent to curve $k(\omega)$ is equal to the reciprocal of the group velocity of light. Indeed, when there is no absorption, the group velocity W, as is well known, satisfies the relationship

$$\frac{1}{W} = \frac{dk}{d\omega} = \frac{1}{c}\frac{d}{d\omega}(\omega n) = \frac{1}{c}\left(n + \omega \frac{dn}{d\omega}\right) \tag{10}$$

Hence, the group velocity of light for frequencies which can be radiated is related to the velocity of motion v and $\cos \Theta$ by the relationships*

$$\frac{v \cos \Theta}{W(\omega)} < 1 \qquad \text{transition from I into II} \tag{11}$$

$$\frac{v \cos \Theta}{W(\omega)} > 1 \qquad \text{transition from II into I} \tag{12}$$

$$\frac{v \cos \Theta}{W(\omega)} = 1 \qquad \text{tangency} \tag{13}$$

At sufficiently large ω, the quantity W becomes equal to c. Indeed, the refraction index tends to unity, and hence curve $k(\omega) = \omega n/c$ approaches a straight line with a slope of $1/c$.

The straight lines a rise more abruptly since $v < c$ and consequently

$$\frac{1}{v \cos \Theta} > \frac{1}{c}$$

Hence, all the three straight lines a at great ω are in Region II.

This entails a number of consequences. First of all, it is evident that the straight line a_1 will necessarily cross curve $k(\omega)$, i.e. Eq. (5), as has already

* The magnitude determined by (10) has the meaning of the group velocity of light only when there is no strong absorption, i.e. in those regions of the spectrum for which the medium is transparent. The part of curve $k(\omega)$ corresponding to the region of anomalous dispersion, in which there is unquestionable dispersion, is shown in Fig. 1 by a dotted line. The peculiarities of radiation for frequencies getting into this region call for special consideration.

been noted, must always have a solution. As a matter of fact, the straight line a_1 passes through point $\omega = \omega_0$ lying on the axis of abscissae, which means that the straight line must go over somewhere from Region I into Region II. Moreover, it means that at any rate a frequency is radiated for which inequality (11) corresponding to a transition from I into II is applicable.

The straight lines a_0 and a_2 as might have been expected, do not always cross curve $k(\omega)$. This requires that their incline to the abscissa axis should be sufficiently small. This means that the velocity should be high and angle Θ should not be large.

At great ω both these straight lines also prove to be in Region II. It follows from this that if there are crossings then at any rate, the last of them which determines the highest of the radiated frequencies corresponds to a transition from Region I into Region II. The result is then again that there is a frequency in the radiation, for which inequality (11) is valid. For forward radiation, i.e. $\Theta = 0$, this means that there is a component for which $v < W$ and, consequently, for at least a part of the radiation, energy is propagated at a higher velocity than that of the source of light[5].

It also follows from the above that if there is a frequency satisfying condition (12) (for instance corresponding to point B_1 on the straight line a_1), the composition of the radiation will infallibly be complex, since there must be a frequency or frequencies satisfying condition (11). (In the general case the number of possible crossings for the straight line a_1 is always odd, and for the straight line a_2 always even.)

The boundary of the appearance of radiation or of new components of radiation is evidently represented by a case when the corresponding straight line a begins to touch curve $k(\omega)$. This means the fulfilment of Eq. (13). With $\Theta = 0$ we obtain, in agreement with the above, $v = W$ for the threshold frequency.

The dotted line a' in Fig. 1 corresponds to the threshold of appearance of the complex effect for the ordinary Doppler effect. As seen from the figure, the frequency begins to split when the slope of the straight line a_1 increases in comparison with that of the dotted line. This means that the complex Doppler effect arises in this case not when the velocity increases in comparison with the threshold velocity, but, quite the contrary, when it decreases or when the angle becomes larger (it is worth recalling that the tangent of the incline of the straight line a_1 equals $1/v \cos \Theta$). This is explained by the fact that the complex Doppler effect takes place here only within some

range of velocities or angles, and the dotted line corresponds to the upper, and not the lower threshold of the effect.

It has been assumed up till now that angle Θ is acute, i.e. that the product $v \cos \Theta$ is positive. What was said above regarding the complex Doppler effect may also be applied to the case of obtuse angles Θ but in this case negative group velocity will have to be taken into consideration. It appears that the threshold for the appearance of the complex Doppler effect, with $\Theta > \pi/2$ is determined by Eq. (13). Quantity $\cos \Theta$ is negative in this case, therefore Eq. (13) is valid only when the quantity W is less than zero. The import of negative group velocity for the Vavilov-Čerenkov effect was first investigated by Pafomov[16,12b] who pointed out that such a case should be real in anisotropic media*. This is a very interesting case. We are accustomed to the idea that the Vavilov-Čerenkov radiation is directed forward at an acute angle. This is, however, correct only if the group velocity is positive. If it is negative, the picture is quite different.

Fig. 2a shows schematically the ordinary case of the Vavilov-Čerenkov radiation. The phase velocity for radiated light $u = c/n$ forms in this case an

* This is related to the fact that in an anisotropic medium the direction of the group velocity does not coincide with that of the phase velocity. This question is treated in the next section of the lecture.

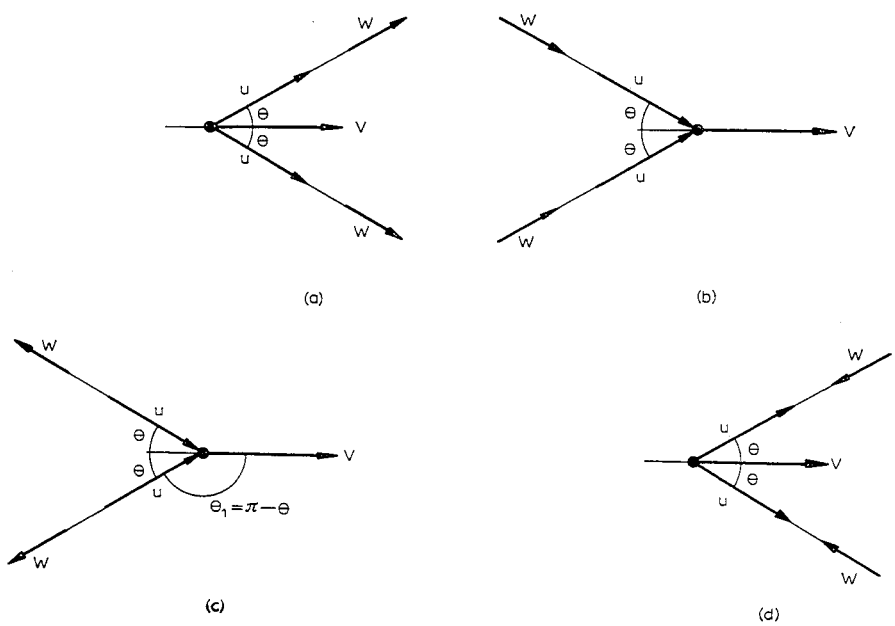


Fig. 2.

acute angle Θ with the direction of velocity v. The equation of electrodynamics also permits of the solution schematically represented in Fig. 2b. The direction of phase velocity, i.e. the direction of wave propagation forms in this case, too, the same acute angle Θ with a velocity vector. The waves do not, however, come from the emitter, but towards it. The first case is interpreted as a radiation of waves, and the second as their absorption. If there is no source of energy feeding the waves, flowing to the emitter, then the case of Fig. 2b is not realizable and the corresponding solution is rejected. But this is correct only if the group velocity is positive, i.e. if its direction coincides with that of phase velocity (see vector W in Figs. 2a and 2b). The direction of the energy flux coincides in this case with the direction of phase velocity and, consequently 2a really corresponds to the radiation of the waves, and 2b to their absorption. In a medium with a negative group velocity, vector W is so directed as to meet vector u (the medium is considered optically isotropic, and hence vectors u and W may be only parallel or anti-parallel). Therefore, with $W < 0$, Fig. 2c corresponds to radiation of energy, and 2d to its absorption. Hence, if the group velocity is negative, the direction of the energy flux of the Vavilov-Čerenkov radiation forms an obtuse angle $\Theta_1 = \pi - \Theta$ with the direction of the velocity, and the motion of the waves is directed not from the particle, but, quite the contrary, towards it.* A similar analysis can also be made of an emitter with a natural frequency ω_0, moving in a medium with a negative group velocity[5,12b].

It can be seen from the above that many substantial peculiarities of radiation in a refractive medium are actually related not only to the magnitude of the phase velocity of light, but also to the group velocity of light. It may be expected that the role of the group velocity of light will reveal itself most distinctly in anisotropic media in which the directions u and W form some angle with one another.

Radiation in optically anisotropic media

Radiation of a light source moving in a crystal should possess a number of features as compared with that in isotropic media. Interest in this range of

* The analysis given in Fig. 2 is similar in many ways to the example given in L. I. Mandelstam's lectures on the refraction of light by a medium with a negative group velocity (*Collected Works* by L. I. Mandelstam, Vol. 5, p. 463).

problems has enhanced recently in connection with the studies of the processes in plasma*. As to propagation of waves, a plasma placed in a magnetic field is similar to a uniaxial gyrotropic crystal.

The Vavilov-Čerenkov effect in crystals was first investigated theoretically by V. L. Ginzburg[13] and then by other investigators (see, for example, review[8]). It has not, however, been studied experimentally to this day.

The equation determining the radiated frequency ω remains the same as in an isotropic medium, i.e., ω is determined by Eq. (4). The magnitude of the refraction index n in the case of an anisotropic medium depends, however, not only on the frequency of light, but also on the angle and polarization. The result is that for the Vavilov-Čerenkov radiation the cone of normals to the wave surfaces is not circular in this case, as in an isotropic medium, but may have quite an odd shape. Thus the direction of velocity does not necessarily coincide with the axis of the cone, and in some cases may even lie beyond the cone[16].

Another peculiarity is related to polarization of the light. The Vavilov-Čerenkov radiation is always polarized. As a rule, polarization of the light in this phenomenon does not attract attention, since it has not been used so far in present-day practical application of the radiation. However, from the viewpoint of the mechanism of the phenomenon, polarization is highly important. It is worth mentioning, for example, that the radiation of a magnetic charge, if it exists at all, could be distinguished at once from the radiation of an electric charge, since in this case the magnetic and electric vectors change places. The question of polarization of light is also of importance for the quite real case of radiation of dipoles and multipoles, though it has not yet been studied experimentally.

The role of polarization is manifested most distinctly in an anisotropic medium. First of all, one can obtain here, depending on the polarization of the radiated light, not one, but two cones of wave normals corresponding to so-called ordinary and extraordinary rays in a uniaxial crystal. Moreover, the distribution of the radiation intensity is a complex function of the angles and is related to polarization of the light. The fulfilment of condition (4) does not suffice to bring about radiation, since the intensity of the waves of a given polarization may prove to equal zero. For example, if a particle moves in the direction of the axis of a uniaxial crystal, the cone of ordinary rays must disappear in the radiation[8].

* Some of the problems connected with plasma are dealt with in I. E. Tamm's Nobel Lecture, see this book, p. 470.

The third peculiarity is related to the fact that in an anisotropic medium the direction of the ray, i.e. the direction of a narrow beam of light, does not, generally speaking, coincide with the normal to the wave surface. There exist such directions of rays in a crystal, for which the normal to the wave surface forms some angle α with the ray (see Fig. 3).

The velocity at which the phase of the wave propagates in the direction of the ray, as can be seen from Fig. 3, is $1/\cos\alpha$ times greater than the phase velocity, i.e. $u' = u/\cos\alpha = c/n\cos\alpha$. We shall call u' the velocity of the waves along the ray. It should not be confused with the group velocity of light, i.e. with the velocity of transfer of light energy which, naturally enough, is also directed along the ray. The group velocity equals velocity u' only under the condition that there is no dispersion of light in the medium. Indeed, the velocity of the waves along the ray does not depend in this case on frequency, and hence the group of waves moves with the same velocity.

The velocity of the waves along the ray is important for radiation in anisotropic media. Let us consider in this connection the threshold velocity for the appearance of the Vavilov-Čerenkov effect. The assertion that the Vavilov-Čerenkov radiation for a light of frequency ω arises at a velocity greater than the phase velocity of light with the given frequency implies

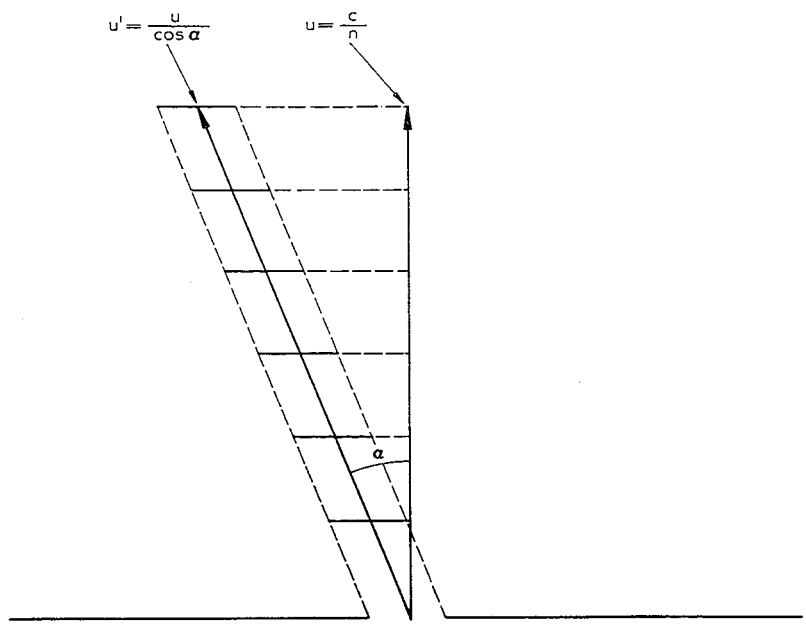


Fig. 3.

that the medium is isotropic. If this statement be considered applicable to anisotropic media (which, as will be seen below, is not always the case), it is necessary, at least, to indicate with which direction of the phase velocity the velocity of motion is to be compared.

Eq. (4), i.e. $(vn/c) \cos \Theta = 1$, is also valid for anisotropic media, and in this case $c/n = u$ is the phase velocity for the given direction of the normal to the wave, forming angle Θ with vector $\vec{v}$. As is well known, when the velocity approaches threshold velocity in an isotropic medium, angle Θ decreases to zero, i.e., the cone of wave normals is compressed in the direction $\vec{v}$. In a crystal, the cone of wave normals is likewise compressed in this case towards some axis which, as a rule does not, however, coincide with $\vec{v}$. If this axis is represented by the direction of the velocity, the threshold $\Theta = 0$ and then we obtain from (4) that $v = c/n$ where c/n is assumed for the direction $\vec{u} = (\overrightarrow{c/n})$ coinciding with $\vec{v}$. Hence, $\vec{v} = \vec{u}$. This relationship actually proves to be correct for boundary velocity in the usual cases of motion in a uniaxial crystal parallel or perpendicular to the optical axis. It has not, however, been stressed that it cannot always be applied.

It may be shown that the general condition for the appearance of the Vavilov-Čerenkov radiation of frequency ω should be formulated in the following way. The threshold velocity of the source of light should equal the velocity of waves along the ray in the direction of motion. In other words, the threshold velocity $\vec{v} = \vec{u}'$. For the threshold velocity, the direction of the ray coincides with $\vec{v}$ and not the normal to the wave which forms an angle α with $\vec{v}$. Hence in the general case, the threshold value is $\Theta = \alpha$.

In a special case, when the direction of the ray coincides with the wave normal in an anisotropic medium, i.e., $\alpha = 0$; $\vec{u}' = \vec{u}$. Then we have $\vec{v} = \vec{u}$ for the threshold velocity. Finally, in an isotropic medium, where the phase velocity of light c/n is the same in all directions, it is possible to go over from vectors to scalar quantities, which means that $v = u$. Hence the well-known statement that the velocity equal to the phase velocity of light is the threshold velocity, has a limited field of application. It is a special case of a more general condition.

The above is easy to explain by using the Huygens principle for plotting the wave surface of radiation. This procedure is still generally used at present to describe the Vavilov-Čerenkov effect in an elementary way, and at the time it was one of the guiding ideas in the creation of its theory. This method can be easily applied to the case of an anisotropic medium.

The Huygens principle is frequently used in crystal optics to explain the

peculiarities of behaviour of the so-called extraordinary ray during the refraction of light. The wave surface is found by the Huygens principle as an envelope of the waves emitted from separate points. Whereas, however, for an isotropic medium a sphere of radius $r = (c/n)\, t$ is plotted around every point, where t is the time of movement of the waves, a crystal calls for a different approach. Of importance is the distance covered by the wave from a given point in the given direction of the ray. The distance equals the velocity of the waves along the ray, multiplied by time t, i.e. $\vec{u}'\, t$. Therefore, the unknown quantity is represented by the envelope of the so-called surfaces of the rays plotted around every source of waves and determined by the equation $\vec{r} = \vec{u}'t$.

Let us apply the Huygens principle to the case of Vavilov-Čerenkov radiation in a uniaxial crystal. The velocity of the ordinary and extraordinary rays is not the same here and, therefore, generally speaking, two cones of waves are obtained. In order not to encumber the drawing, they are shown on separate Figs. 4 and 5. We have to consider each point of the particle trajectory as a source of waves. In this case the wave phase is determined by the instant of passage of the particle through a given point. Let us assume that at moment $t = -t_3$ the emitter was at point A_3, at moment $t = -t_2$ at point A_2, at moment $t = -t_1$ at A_1, and, finally, at the moment of observation $t = 0$ at point A_0.

For ordinary rays, the velocity of the waves along the ray, as in an isotropic medium, is equal to the phase velocity of light c/n and does not depend on the direction. The surfaces of the rays are simply spheres whose radii for points A_3, A_2, A_1, and A_0 are $(c/n)\, t_3$, $(c/n)\, t_2$, $(c/n)\, t_1$ and 0 respectively (see Fig. 4). The envelope of these spheres evidently represents a cone of

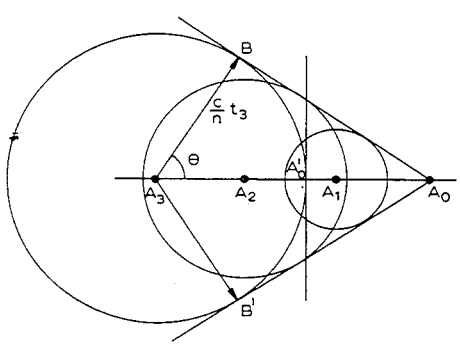


Fig. 4.

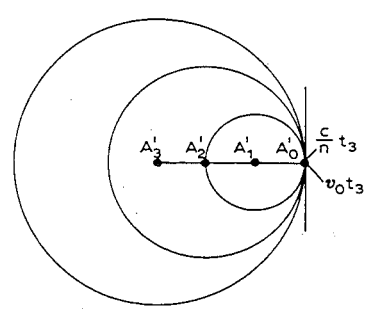


Fig. 4a.

circular cross section with the apex at A_0 *. Its generatrices lying in the plane of the drawing are A_0B and A_0B'.

According to the Huygens principle, the directions of the rays are defined by the radius vectors drawn from some centre of the waves to the point of tangency with the envelope. For example, in Fig. 4 it is A_3B or A_3B' coinciding with the generatrices of the wave normal cone for ordinary rays. Thus the radiation cone is obtained for ordinary rays in the same way as in the Vavilov-Čerenkov effect in an isotropic medium. The substantial difference from an isotropic medium is related to the polarization of light and the distribution of intensity, depending on it. This was not taken into account in the construction.

From Fig. 4 it is not difficult to determine the magnitude of the threshold velocity. When the velocity diminishes, the distances between points A decrease. The threshold case arises when point A_0 occupies the position of A_0' on the surface of the sphere. (This case is depicted separately in Fig. 4a.) At lower velocities, one of the spheres lies completely within the other and they do not have a common envelope. In the threshold case, they have only a common point of tangency A_0'. Thus evidently $(c/n)\ t_3 = v_0 t_3$, i.e. $v_0 = c/n$. The cone of wave normals is compressed in the direction of velocity v, and the wave cone transforms into a plane perpendicular to the axis of motion at point A_0' (Fig. 4a).

The Huygens principle can also be applied in a similar way to obtain a wave cone for the extraordinary rays (Fig. 5). The difference consists in that surfaces of rays $\vec{u}'t_3$, $\vec{u}'t_2$, and $\vec{u}'t$, instead of spheres are plotted around points A_3, A_2, and A_1. The cone enveloping the surfaces with an apex at A_0 is not circular in the case shown in Fig. 5. The generatrices of this wave cone A_0C and A_0C' lie in the plane of the drawing. The lines perpendicular to them, for instance A_3D and A_3D' determine the wave normals, and their length is proportional to the phase velocities. The vectors drawn from A_3 to the points of tangency A_3F and A_3F' indicate the corresponding directions

* Strictly speaking, such an analysis presupposes that there is a superposition of monochromatic waves. Each point of the trajectory should, therefore, be regarded as a source of such waves emitted for an infinitely long time. Actually, it is only the summation of waves of various frequency that produces a light impulse when the particle passes through a given point. Hence, there exists, of course, not one, but an unlimited multitude of wave surfaces for a given frequency. The one that is generally plotted is singled out only by its passage through the instantaneous position of the particle (which we shall term the wave cone).

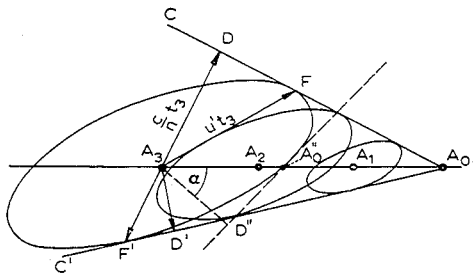


Fig. 5.

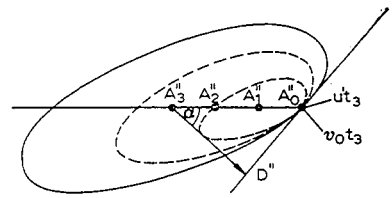


Fig. 5a.

of rays, which, as seen from Fig. 5, do not coincide with the wave normals. It can also be seen from the drawing that the direction of an extraordinary ray for the Vavilov-Čerenkov radiation in a crystal may even constitute an obtuse angle with the direction of velocity (direction A_3F' in Fig. 5).

It is not difficult to determine the magnitude of the threshold velocity for the appearance of extraordinary rays in the Vavilov-Čerenkov radiation, which, generally speaking, differs from threshold velocity for ordinary rays. The threshold case occurs when the velocity diminishes to such an extent that point A_0 coincides with point A_0''. In this case all the surfaces of the rays lie within each other and have a common point of tangency A_0''. It can be seen from Fig. 5 or 5a showing a threshold case that the threshold value is $v = v_0 = u'$. The wave cone then transforms into plane $A_0''D''$ and the wave normal forms an angle α with direction v. By tracing what happens to the cone of wave normals (its generatrices are A_3D and A_3D' in Fig. 5) during a decrease in velocity, i.e., when point A_0 approaches A_0'', it is not difficult to prove that it is compressed not in direction v but in direction A_3D''. Hence, in a threshold case in Eq. (4), it may be assumed that not $\Theta = 0$ but $\Theta = \alpha$. Then Eq. (4) produces $(vn/c)\cos\alpha = 1$, i.e., actually $v = c/n\cos\alpha = u'$.

It is worth recalling that with the aid of Figs. 4 and 5 we have determined the threshold of appearance of light of some given frequency ω. The velocity at which radiation generally appears is determined by a minimal magnitude of wave velocity of waves along the ray, namely $u' = u'_{min}$ in a given medium for a ray directed along the motion. For frequency ω' for which $u' = u'_{min}$ the velocity of the waves along the ray does not depend on frequency and is thus equal to the group velocity. Hence, we again come to the conclusion that the threshold is related to the group velocity.

The analysis of radiation of a system possessing a natural frequency of oscillations ω_o may also be applied to the case of an optically anisotropic medium. The same peculiarities are manifested here as referred to in connection with the Vavilov-Čerenkov radiation. The connection between ω, Θ, v, and ω_o is determined, as before, by the same Eqs. (5) and (6) as in an isotropic medium, but now quantity n refers to the direction of a wave normal at an angle Θ to the velocity.

The dependence of n on the direction leads to the fact that the connection between Θ and the frequency of radiation ω at preset natural frequency ω_o and velocity v is not elementary. To find Θ, use can be made of the graphic method suggested by V. E. Pafomov[16] for analysing the Vavilov-Čerenkov effect in crystals, applying it to the case of an arbitrary ω_o (see Fig. 6). The figure shows a section of a surface of wave vectors $\vec{k}(\omega) = (\overrightarrow{\omega n})/c$ for the given ω in the case of extraordinary rays in a uniaxial crystal. The surface indicating the dependence on the direction of vectors $\vec{k}$ (they are oriented along the normal to the wave) differs from that of refraction indices only by a constant factor ω/c (we consider magnitude ω as prescribed). Thus, for

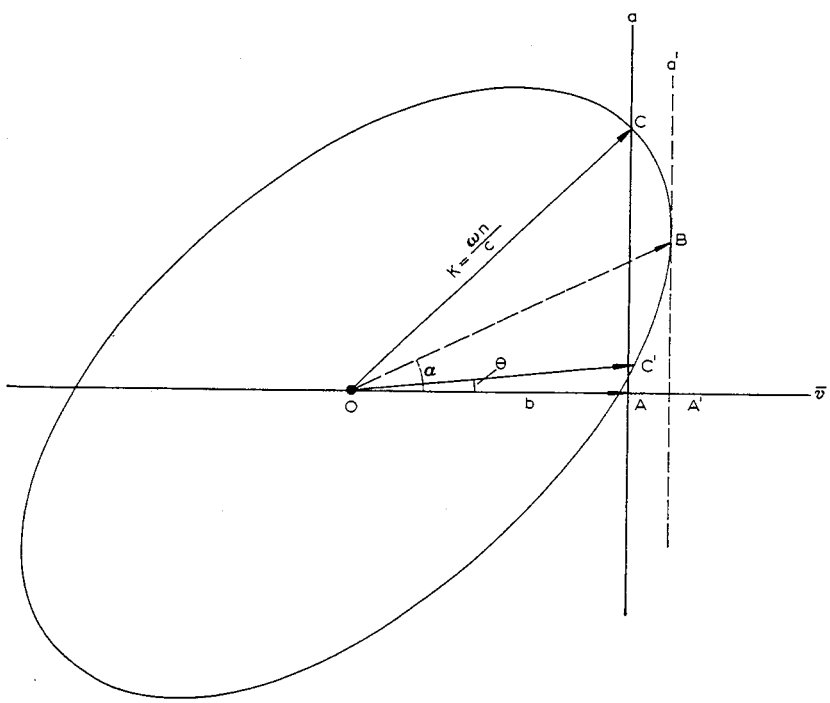


Fig. 6.

a uniaxial crystal, the surface represents an ellipsoid of rotation. Let us assume that axis $\vec{v}$ is the direction of motion of the emitter. Let us plot on axis $\vec{v}$ segment OA of length b which equals b_0, b_1, or b_2 depending on whether the analysis deals with the Vavilov-Čerenkov effect, the Doppler ordinary effect, or the Doppler super-light effect. Then

$$b_0 = \frac{\omega}{v} \tag{14}$$

$$b_1 = \frac{\omega - \omega_0}{v} \tag{15}$$

$$b_2 = \frac{\omega + \omega_0}{v} \tag{16}$$

At point A which is the end of b we shall plot plane a perpendicular to axis v. Let us consider the curve where the plane crosses surface $k(\omega)$ as a section of some cone with the apex at O. The generatrices of this cone OC and OC′ lie in the plane of the figure. The cone defines the magnitude and direction of vectors $\vec{k}$ for light of frequency ω appearing in the case under consideration, i.e. for the given kind of radiation with preset ω_0 and v.

Indeed, as can be seen from Fig. 6, OA $= b$ is a projection of vector OC or OC′, i.e. of vector $k = \omega n(\omega, \Theta)/c$. Hence

$$\frac{\omega n(\omega, \Theta)}{c} \cos \Theta = b \tag{17}$$

By substituting the values of b from (14), (15), or (16) we obtain identical equations (4), (5), or (6).

It can be seen from Fig. 6 that the cone of wave normals may actually be not only asymmetric, but, as has already been mentioned, axis $\vec{v}$ may even lie outside the cone.

Plane a does not always cross the surface of $k(\omega)$. This corresponds to the evident fact that not every frequency is radiated for given v and ω_0. If $b = b' =$ OA′ (see Fig. 6), the plane touches the surface and, consequently, $b' =$ OA′ is a boundary for the appearance of the given frequency ω in the spectrum. Vector $\vec{k}$, i.e. the wave normal, coincides in this case with OB. It can be easily proved that it forms angle α with the direction of velocity, and the direction of the ray coinciding with that of motion. If, in accordance

with this, angle $\Theta = \alpha$ is inserted in Eq. (3), we obtain the following general condition for velocity v_o required for the appearance of frequency ω

$$\frac{\omega}{u'} = \frac{\omega \pm \omega_o}{v_o} \tag{18}$$

where u' is the velocity of the waves along axis $\vec{v}$ (positive or negative, i.e. directed along $\vec{v}$ or opposite to it). In a special case of the Vavilov-Čerenkov radiation, $\omega_o = 0$.

Radiation of a system possessing a natural frequency of oscillations and moving in an optically anisotropic medium was first studied by K. A. Barsukov and A. A. Kolomensky[17]. They elucidated a number of peculiarities of radiation related to the presence of ordinary and extraordinary rays and the significant role of wave polarization.

It is highly interesting that this seemingly more complex case appears to present even now some interest from an experimental point of view. Barsukov and Kolomensky made a special study of radiation of radio waves in the ionosphere which behaves like an optically anisotropic medium under the action of the earth's magnetic field. It is important that this medium possesses strong dispersion at some range of frequencies and that the complex Doppler effect is possible in it. Kolomensky and Barsukov have pointed out that this phenomenon may take place in the case of radio waves of appropriate frequency, transmitted by an artificial earth satellite moving in the ionosphere. They found that the Doppler shift of frequency of the order of ten to a hundred cycles per second should be accompanied in this case by splitting of the radiated frequency into components of several hundredths of a cycle per second apart. Apparently, with a well-stabilized frequency of the transmitter, such splitting could be detected.

I have aimed to prove in my lecture that there is a wide range of problems related to the radiation of sources of light, moving in refractive media. Radiation of an electric charge moving at super-light velocity in an isotropic medium, i.e. the experimentally investigated case of the Vavilov-Čerenkov effect, is, in essence, but a special, though a highly interesting instance in this realm of phenomena.

1. P. A. Čerenkov, *Dokl. Akad. Nauk SSSR*, 2 (1934) 451.
S. I. Vavilov, *Dokl. Akad. Nauk SSSR*, 2 (1934) 457.
P. A. Čerenkov, *Tr. Fiz. Inst. Akad. Nauk SSSR im. P. N. Lebedeva*, 2, No. 4. (1944)
2. I. E. Tamm and I. M. Frank, *Dokl. Akad. Nauk SSSR*, 14 (1937) 107.
3. V. L. Ginzburg, *Zh. Eksperim. i Teor. Fiz.*, 10 (1940) 589.
4. I. M. Frank, *Izv. Akad. Nauk SSSR, Ser. Fiz.*, 6 (1942) 3; *J. Phys. USSR*, 7 (1943) 49.
5. I. M. Frank, *Zh. Eksperim. i Teor. Fiz.*, 36 (1959) 823; *Joint Inst. Nucl. Res.*, P-237 (1958).
6. E. Fermi, *Phys. Rev.*, 57 (1940) 485.
7. I. M. Frank, *Vavilov Memorial Symposium*, USSR Academy of Sciences Publishing House, Moscow, 1952.
8. B. M. Bolotovsky, *Usp. Fiz. Nauk*, 62 (1957) 201.
9. V. L. Ginzburg and I. M. Frank, *Dokl. Akad. Nauk SSSR*, 56 (1947) 699.
10. V. L. Ginzburg and V. Y. Eidman, *Zh. Eksperim. i Teor. Fiz.*, 35 (1958) 1509.
L. S. Bogdankevich, *Zh. Tekhn. Fiz.*, 29 (1959) 1086.
11. V. L. Ginzburg and I. M. Frank, *Zh. Eksperim. i Teor. Fiz.*, 16 (1946) 15; *J. Phys. USSR*, 9 (1945) 353.
12. G. M. Garibyan, *Zh. Eksperim. i Teor. Fiz.*, 33 (1957) 1403.
V. E. Pafomov, *Zh. Eksperim. i Teor. Fiz.*, 36 (1959) 6.
13. V. L. Ginzburg, *Zh. Eksperim. i Teor. Fiz.*, 10 (1940) 608; *J. Phys. USSR*, 3 (1940) 101.
14. V. L. Ginzburg and I. M. Frank, *Dokl. Akad. Nauk SSSR*, 56 (1947) 583.
15. I. M. Frank, *Usp. Fiz. Nauk*, 30 (1946) 149.
16. V. E. Pafomov, *Thesis, Physical Insitute of the USSR Academy of Sciences, Moscow*, 1957; *Zh. Eksperim. i Teor. Fiz.*, 32 (1957) 366; 37 (1959) No. 6.
17. K. A. Barsukov and A. A. Kolomensky, *Zh. Tekhn. Fiz.*, 29 (1959) 954.
18. J. V. Jelley, *Čerenkov Radiation and its Applications*, Pergamon Press, 1958.

Biography

Il'ja Michajlovič Frank was born in Leningrad on October 23, 1908, the younger son of Mikhail Lyudvigovič Frank, a Professor of Mathematics, and his wife, Dr. Yelizaveta Mikhailovna Gratsianova. He attended the Moscow State University as a pupil of Vavilov, and graduated in 1930. In 1931 he became a senior scientific officer in Professor A. N. Terenin's laboratory in the State Optical Institute in Leningrad, and in 1934 he joined the P. N. Lebedev Institute of Physics of the U.S.S.R. Academy of Sciences as a scientific officer. He was promoted firstly to senior scientific officer and, in 1941, to his present position as officer in charge of the Atomic Nucleus Laboratory. Since 1957 he has simultaneously occupied the post of Director of the Neutron Laboratory of the Joint Institute of Nuclear Investigations.

The first investigations of I. M. Frank were in the field of photoluminescence and in photochemistry. From 1934 he began his work on nuclear physics in the Laboratory of Professor D. V. Skobeltzyn. The experimental investigations of pair creation by γ-rays and other problems connected with the measurements and application of γ-rays were carried out by him. His further works were devoted to neutron physics, the investigation of reactions on light nuclei and nuclear fission by mesons.

The subject of his theoretical investigations is the Vavilov-Čerenkov effect and related problems.

Frank was awarded the degree of Doctor of Physico-Mathematical Sciences in 1935; in 1944 he was confirmed in the academic rank of Professor, and was elected a Corresponding Member of the U.S.S.R. Academy of Sciences in 1946.

He married Ella Abramovna Beilikhis, a noted historian, in 1937. They have one son, Alexander.

IGOR' E. TAMM

General characteristics of radiations emitted by systems moving with super-light velocities with some applications to plasma physics

Nobel Lecture, December 11, 1958

The mechanism of radiation of light by a system moving with a super-light velocity is a very simple one and common to the radiation at corresponding conditions of all kinds of waves – electromagnetic as well as sound waves, waves on the surface of water, etc.

Consider a system which in principle is able to emit the radiation in question – e.g. an electrically charged particle in the case of light, a projectile or an airplane in the case of sound, etc. As long as the velocity of this system as a whole is smaller than the velocity of propagation of waves in the surrounding medium, the radiation can be produced only by some oscillatory motion of the system or of some of its parts – e.g. by the oscillation of an electron in an atom or by the revolutions of the propellers of a plane. The frequency of the radiation emitted is evidently determined by the frequency of the oscillations in question. To be more exact, for the radiation to be possible the motion has not necessarily to be a periodic one, but it has to be non-uniform* (i.e. its velocity should not be constant in time).

But when a velocity of the system becomes greater than that of the waves in question, quite a new mechanism of radiation is introduced, by means of which even systems possessing a constant velocity radiate. Let $c'(\omega)$ denote the velocity of propagation in the surrounding medium of waves, possessing the frequency ω. Then as a rule the radiation of a system moving in the medium with a constant velocity v, embraces all the frequencies which satisfy the fundamental condition

$$v > c'(\omega) \qquad (1)$$

* About an exception to this rule – the so-called transition radiation – see V. L. Ginzburg and I. Frank[1] (1945).

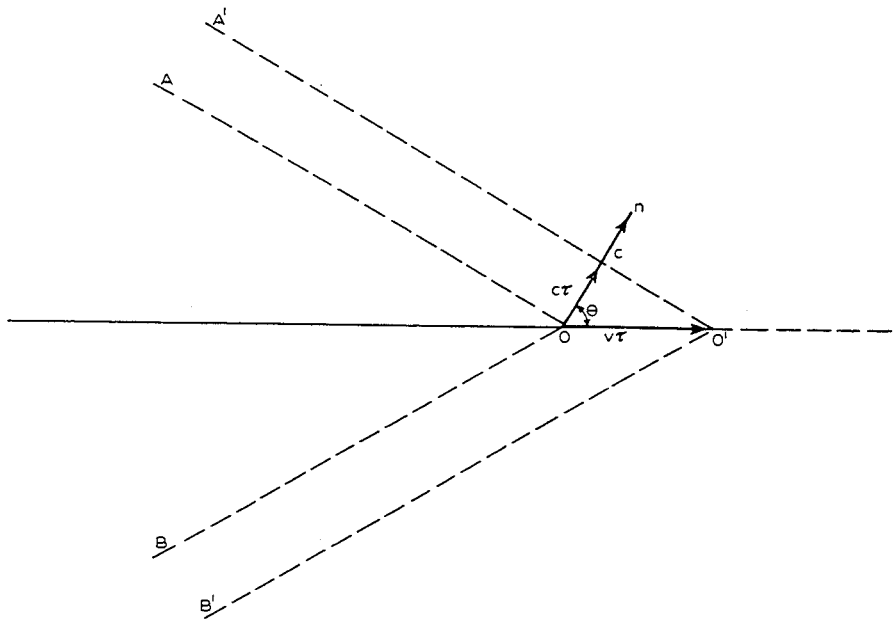


Fig. 1.

This radiation is characteristically a very directional one – waves of a given frequency ω are emitted only under a definite angle Θ to the direction of motion of the system, this angle being determined by the relation

$$\cos \Theta = \frac{c'(\omega)}{v} \tag{2}$$

To prove these fundamental relations one has only to take account of the fact that at all velocities, whether small or large, the field of a uniformly moving system must be stationary with respect to this system. If the system radiates, it means that in its field at least one free wave is present (a free wave of a frequency ω is by definition propagated in the medium with the characteristic phase velocity $c'(\omega)$ to any distance, however far from the source of the wave). Let O and O′ (Fig. 1) be the positions of the uniformly moving system at two consecutive moments $t = 0$ and $t = \tau$. The phase of the wave radiated by the system must be stationary with respect to the system. It means, that if AO is that front of the wave* which at the moment

* The fronts of the wave are conical, due to the cylindrical symmetry; AOB is the projection on the plane of drawing of such a cone.

$t = 0$ passes through the system at O, then this front, being propagated in the medium with the velocity $c'(\omega)$, will permanently keep up with the system, and in particular will at the moment $t = \tau$ occupy such a position AO′, as to pass through O′. Now the direction $\vec{n}$ of propagation of a free wave is perpendicular to its front, therefore the triangle OCO′ is a rectangular one and we easily obtain from it the fundamental relation (2).

Since the value of a cosine cannot exceed unity, Eq. (1) follows directly from (2).

All these general properties of the radiation in question were for a very long time well known in aerodynamics. The air waves emitted at supersonic velocities are called Mach waves. The emission of these waves sets in when the velocity of a projectile or of a plane begins to exceed the velocity of sound in the air. Emitting waves means losing energy and these losses are so large that they constitute the main source of resistance to the flight of a supersonic plane.

That is why in order to cross the sound barrier, i.e. to achieve supersonic velocities in aviation, it was necessary to increase very substantially the power of the engines of a plane.

We perceive the Mach waves radiated by a projectile as its familiar hissing or roaring. That is why, having understood the quite similar mechanism of the Vavilov-Čerenkov radiation of light by fast electrons, we have nicknamed it «the singing electrons».

I should perhaps explain that we in the USSR use the name «Vavilov-Čerenkov radiation» instead of just «Čerenkov radiation» in order to emphasize the decisive role of the late Prof. S. Vavilov in the discovery of this radiation.

You see that the mechanism of this radiation is extremely simple. The phenomenon could have been easily predicted on the basis of classical electrodynamics many decades before its actual discovery. Why then was this discovery so much delayed? I think that we have here an instructive example of a situation not uncommon in science, the progress of which is often hampered by an uncritical application of inherently sound physical principles to phenomena, lying outside of the range of validity of these principles.

For many decades all young physicists were taught that light (and electromagnetic waves in general) can be produced only by *non-uniform* motions of electric charges. When proving this theorem one has – whether explicitly or implicitly – to make use of the fact, that super-light velocities are forbidden

by the theory of relativity (according to this theory no material body can ever even attain the velocity of light). Still, for a very long time the theorem was considered to have an unrestricted validity.

So much so, that I. Frank and I, even after having worked out a mathematically correct theory of Vavilov-Čerenkov radiation, tried in some at present incomprehensible way to reconcile it with the maxim about the indispensibility of acceleration of charges. And only on the very next day after our first talk on our theory in the Colloquium of our Institute we perceived the simple truth: the limiting velocity for material bodies is the velocity of light *in vacuo* (denoted by c) whereas a charge, moving *in a medium* with a constant velocity v, will radiate under the condition $v > c'(\omega)$, the quantity $c'(\omega)$ depending on the properties of the medium. If $c'(\omega) < c$, then this condition may very well be realized without violating the theory of relativity ($c' < v < c$).

When we first discussed our theory with Professor A. Joffe he pointed out to us that A. Sommerfeld[2] as long ago as 1904 has published a paper, dealing with the field of an electron possessing a constant velocity greater than that of light, and has calculated the resistance to such a motion, due to the radiation, emitted by the electron. But Sommerfeld considered only the motion of an electron *in vacuo*. A year later the theory of relativity came into existence, the motion considered by Sommerfeld was proved to be impossible, Sommerfeld's paper was completely forgotten and for the first time in many years was referred to in our paper[3] of the year 1937.

Let us return now to general characteristics of radiation emitted at superlight velocities. In addition to those already indicated a new and very peculiar one emerged in papers of I. Frank[4] in 1943 and of V. L. Ginzburg and I. Frank[5] of 1947.

Suppose that a system A, moving with a constant velocity $\vec{v}$, radiates an amount of energy ε in a direction characterized by a unit vector $\vec{n}$. The balance of energy gives the relation

$$\varepsilon + \Delta T + \Delta U = 0 \tag{3}$$

where ΔT and ΔU denote respectively the increase, caused by the radiation, of the kinetic energy T of the translational motion of the system A and of the energy U of its internal degrees of freedom. On the other hand, if the radiated energy ε is propagated in the medium with the velocity c'

in a definite direction $\vec{n}$, it necessarily possesses a momentum* ε/c', directed along $\vec{n}$. Therefore the conservation of momentum leads to the vector equation

$$(\varepsilon/c')\vec{n} + \Delta\vec{p} = 0 \tag{4}$$

where $\vec{p}$ is the momentum of the system A. If the increase $\Delta\vec{p}$ of $\vec{p}$ is small in relation to $\vec{p}$, then, according to a general rule,

$$\vec{v}\cdot\Delta\vec{p} = \Delta T \tag{5}$$

Combining these simple and general relations one gets

$$\Delta U = -\varepsilon\left(1 - \frac{v\cos\Theta}{c'}\right) \tag{6}$$

where Θ is the angle between $\vec{v}$ and $\vec{n}$.

If the system A possesses no internal degrees of freedom (e.g. a point charge), then $\Delta U = 0$ and Eq. (6) reduces to the already discussed Eq. (2). Thus we have obtained this fundamental equation once again, but by a new way of reasoning. On the other hand, if the system possesses internal (say, oscillatory) degrees of freedom, and if its velocity is small ($v \ll c'$), then, as usual, the internal energy U of the system decreases by an amount equal to the amount ε of the energy radiated.

But at super-light velocities ($v > c'$) the value of the bracket in (6) may become negative, so that radiation of energy by the system may be accompanied by a *positive* increase ($\Delta U > 0$) of its internal energy U. For example, an atom, being originally in the stable state, radiates light and at the same time becomes excited! In such a case the energy both of the radiation and of the excitation is evidently borrowed from the kinetic energy T, i.e. the self-excitation of a system is accompanied by a corresponding slowing down of the motion of this system as a whole.

* For the case of electromagnetic radiation it was shown first by quantum-theoretical reasoning (Ginzburg, 1940) and then by means of classical electrodynamics (Marx and Györgyi[6], 1955) that ε/c' (c' being the phase velocity) is in fact equal to the total momentum radiated, which comprises both the momentum of the radiation proper and the momentum acquired by the medium.

The relation (6) emerged in discussion of optical problems but it is of a quite general nature and it may turn out to be useful to apply it in aerodynamics (just as Mach's aerodynamical relations (1) and (2) turned out to be useful in optics).

Certainly, a correct calculation of a supersonic motion will automatically take in account everything, including the possible self-excitation of some particular modes of vibrations of a supersonic plane. However, such calculations are necessarily extremely complicated, so that the relation (6) may prove to be useful in giving an insight in the general mechanism of some of the phenomena which become possible at supersonic velocities. On the other hand, Eq. (6) takes in account only the radiative damping of oscillations, whereas in the case of mechanical vibrations of a plane this kind of damping is under ordinary conditions quite negligible in comparison with the damping caused by the internal friction in the vibrating materials. In short, we must consider it an open question whether the phenomena indicated may be of any importance in the complicated problem of a supersonic flight.

Let us now consider as an example some applications of the general theory to a special field, namely to plasma physics.

In a preparatory way we begin with some remarks on the mechanism of energy losses, experienced by fast charged particles travelling through matter. Vavilov-Čerenkov's radiation accounts only for a part – and usually a very small part – of these losses, which are largely due to the ionization and excitation of the medium traversed by the particles. However the mathematical treatment, used by Frank and myself to calculate the radiation losses, proved to be useful for the general problem also and was extended in 1940 by Fermi[7] so as to cover the total energy loss of a charged particle, with the exception of the losses caused by head-on collisions of the particle with atoms of the medium. The losses of the later kind must be calculated separately. The main difference between Fermi's work and ours is that we assumed the medium traversed by the particle to be transparent, whereas Fermi took in account not only the polarization of the medium by the electrical field of the particle, as we did, but also the absorption of electromagnetic waves in it. Fermi has shown, that the screening of the field of the particle, which is caused by the polarization of the medium, and which was not taken in account in previous work on this subject, very considerably reduces the energy losses of very fast particles.

We will not review here the very extensive work on the subject, in which

Fermi's theory was further elaborated and extended. But to obtain some insight into the underlying mechanism we will consider in some detail the processes taking place in a plasma (e.g. a highly ionized gas), which for our purposes may be considered as the simplest of all media. I have myself not done any work on this subject, so that I will report on the work of others, mentioning by name the authors only of relatively new papers, without explicit references to classical works such as e.g. by N. Bohr.

Energy losses of a charged particle traversing plasma can be divided in two parts. Imagine a cylinder of a radius equal to the Debye's radius $D = (\varkappa T/4\pi Ne^2)^{\frac{1}{2}}$, the axis of the cylinder coinciding with the path of the particle. The interaction of the particle considered with plasma particles lying *inside* the cylinder must be treated microscopically; resulting energy losses will be referred to as those due to close collisions. But the interaction of the particle considered with the plasma lying *outside* the cylinder can be treated macroscopically; resulting energy losses will be designated as coherent ones. Under ordinary conditions losses of both kinds are of about equal importance, but in a very hot and rarefied plasma, so important in thermonuclear research, the cross section for the direct Coulomb interaction of charged particles decreases and the coherent losses eventually become preponderant.

Since the index of refraction n of a plasma is for all frequencies less than 1, so that the velocity of light $c' = c/n$ in plasma is greater than its velocity c in vacuo, it may appear that the Vavilov-Čerenkov effect should be absent in plasma. But that is not the case. Firstly, only the velocity $c'(\omega)$ of transverse electromagnetic waves in a plasma exceeds c *at all frequencies*, but not so the velocities of plasma waves proper. Those are longitudinal waves, in which oppositely charged plasma particles oscillate in opposite directions, the restoring force being provided by the resulting electric field. Secondly, in a magnetic plasma, i.e. in a plasma exposed to an external magnetic field, both kinds of waves become interconnected, so that no sharp distinction can be drawn between the transverse and the longitudinal waves. As a result the index of refraction of light varies with the directions of its propagation and polarization, and in a certain range of these directions becomes greater than 1, so that the Vavilov-Čerenkov effect becomes possible.

Let us first consider coherent energy losses of a charged particle moving in a plasma in the absence of external magnetic fields. Almost all these coherent losses are due to the excitation of longitudinal plasma waves by a mechanism

equivalent to the mechanism of Vavilov-Čerenkov radiation of light. To be more exact the phase velocity of plasma waves is equal to

$$c' = \sqrt{3v_T^2 + \frac{\omega_o^2}{k^2}}$$

where $k = 2\pi/\lambda$ is the wave vector,

$$\omega_o = \left(\frac{4\pi Ne^2}{m}\right)^{\frac{1}{2}}$$

is the so-called plasma frequency and v_T is the mean thermal velocity of plasma electrons. As long as the velocity v of the particle considered is less than $\sqrt{3}v_T$, the necessary condition $v > c'$ for the emission of plasma waves cannot be satisfied; and therefore practically all energy losses experienced by the particle are due to close collisions. But when v exceeds $\sqrt{3}v_T$ the condition $v > c'$ is satisfied for a certain range of wavelengths $\lambda = 2\pi/k$ and the coherent losses are switched in*.

Allow me now to make a digression and to turn your attention from plasma to solid metals. At high enough frequencies the valence electrons in a metal can be considered as free and thus as forming together with the atom cores a kind of plasma. The plasma frequency ω_o is proportional to the square root of the density of plasma electrons. Since this density is in a metal far greater than in an ordinary plasma, the frequency of plasma waves in metals is rather high, of the order of $\hbar\omega \sim 10$ eV.

In analogy to the case of an ordinary plasma we have to expect that a fast electron traversing a metal foil will experience, besides other kinds of energy losses, also losses due to the excitation of plasma waves by the mechanism just described. Now that is in fact the case. It is well known that fast electrons traversing a thin metal foil often experience in it large discrete energy losses of the order of 10 eV. I refer you to a comprehensive article by D. Pines[8] (1956), where it is shown that an elementary theory of the plasma excitation in a metal by a fast charged particle, very similar to the theory outlined above for the case of an ordinary plasma, fits the experimental facts relating to discrete energy losses in metals so well, that, in words of the author: « What puzzles exist have to do with why the

* The fact that long plasma waves are very strongly absorbed in plasma itself has no influence on the phenomenon, since the condition of radiation $c'(\omega) < v$ is satisfied only for short enough plasma waves ($\lambda < D$), the damping coefficient of which is small in comparison with their frequency.

agreement is so good, rather than with explaining existing disagreements.»

Turning again to ordinary plasma I would like to emphasize, that the absorption of plasma waves in the plasma itself is conditioned by a reverse Vavilov-Čerenkov effect.

Ordinarily the necessary condition for a marked absorption of waves is the existence of a resonance between the frequency of the wave and a frequency of the absorbing system, e.g. an atom. Thus a free electron, which in distinction to a bound electron possesses no eigen-frequency, performs in the field of a wave periodic oscillations, alternatively acquiring and again losing kinetic energy and thus producing no substantial absorption.

But there exists also another non-resonant mechanism of absorption. If the velocity v of a free electron is greater than that of the wave ($v > c'$), then the projection of the velocity of the electron on the direction of propagation of the wave $v \cos \Theta$ may become equal to the velocity of the wave:

$$v \cos \Theta = c' \tag{7}$$

In this case the electron so to say rides on the crest of the wave, being exposed to a force, the direction of which does not alter in time, and thus continually absorbs energy from the wave until its velocity increases so much, that it drops out of phase with the wave.

Such is the mechanism of absorption of plasma*; the condition (7), which sorts out those plasma electrons which take part in the process of absorption, is identical with the fundamental condition (2) for radiation**.

The damping coefficient γ of plasma waves was first calculated by Landau[9] in 1946. Changing the notations used by Landau one can present the exponential term in Landau's formula in the following form

$$\gamma \backsim \exp\left(-\frac{mu^2}{2\varkappa T}\right) \tag{8}$$

* In principle this mechanism of absorption was indicated as long ago as 1949 by Bohm and Gross[11]. The work of these authors is intimately connected with earlier work of A. Vlasov. A detailed and a very lucid mathematical treatment of this subject was presented by R. Z. Sagdeev and V. D. Shafranov at the Geneva Atoms for Peace Conference last September.

** Radiation takes place if there is say one electron of velocity $\vec{v}$ or a cluster of such electrons, the dimensions of the cluster being small in comparison with the length of the wave radiated. If however electrons of a given velocity $\vec{v}$ are distributed continuously in space, then they do not radiate, since their wave-fields are destroyed by mutual interference. But they do absorb.

where $u = \omega_0/k$. In the range of validity of Landau's formula ω_0/k equals the velocity c' of the wave in question.

Therefore according to (8) the damping of a plasma wave is proportional to the density of plasma electrons, possessing according to Maxwell's law a velocity u, equal to the velocity of the wave. This is in exact correspondence to the mechanism of absorption just indicated.

In a recent paper on the mechanism of the sporadic solar radio-emission Ginzburg and Zhelesniakov[10] (1958) applied and extended the theory outlined above to a new and very interesting domain of physics, the foundations of which were laid in Sweden by Professor Alfvén. In particular they have shown that the known instability of a beam of charged particles traversing plasma, is from a quantum theoretical point of view due to the negative absorption of plasma waves by the beam of particles (the *induced* radiation of waves by the beam particles prevailing over the true absorption).

Before finishing I would like to mention one problem, which plays a rather important role in the present fascinating world-wide effort to harness thermonuclear reactions for peaceful uses – the problem how to heat the plasma. First stages of heating can be easily achieved by exciting an electric current in the plasma. However, the cross-section for Coulomb collisions of charged particles decreases inversely to the fourth power of their relative velocities and in a hot and rarefied plasma these collisions become so rare as to become negligible. Evidently heating by electric currents thus becomes impracticable: only a very small part of the energy of the ordered motion of plasma electrons, excited by an external field, is under these conditions converted into Joule heat.

Many different methods to achieve further heating of the plasma are now being discussed, e.g. the so-called magnetic pumping. I wish to make some remarks on only two such methods, intimately connected with our subject.

First, the heating by a beam of fast charged particles, injected into plasma from outside, is in principle feasible even if the plasma is hot and rarefied. Although in such a plasma energy losses of fast particles due to close collisions become negligible, coherent energy losses, described earlier, are independent of the collision cross-section and become all-important.

It is necessary to stress in this connection two points. First, the heating can in principle be achieved by a beam of fast charged particles travelling not in the plasma itself, but outside it and parallel to its surface. In fact, as we have seen, coherent energy losses are due to the emission of plasma waves

by the fast particles. Now, those of these waves, the length of which is large in comparison with the distance of the beam from the surface of the plasma, will be excited by an external beam much to the same degree as by a beam traversing plasma. The possibilities offered by an external beam were first pointed out by L. Mandelstam for the case of the ordinary Vavilov-Čerenkov radiation. Later Ginzburg[12] (1947) proposed a method of generating micro-radiowaves by means of fast particles travelling along the surface of an appropriate dielectric or in a tunnel bored through the dielectric.

The second point is that if the beam consists of a succession of separate clusters of charged particles, then all the particles of each cluster will generate coherently those of the plasma waves, the length of which is large in comparison with the dimensions of the clusters. Therefore the intensity of these waves will be proportional not to the number of particles in a cluster, but to the square of this number. Evidently this offers the possibility of enhancing the radiation and the heating effect of a beam very considerably.

Let us now turn to another possible method of heating. Morozov[13] (1958) has recently calculated the excitation of so-called magneto-acoustic waves in a magnetic plasma (i.e. a plasma exposed to a constant external magnetic field) by an electric ring-current, moving with a sufficient velocity in a direction perpendicular to the plane of the ring-current. The current may move within the plasma – one can imagine a plasma ring, bearing a current, the ring being injected from outside into the plasma to be heated. Otherwise the current in question may be flowing outside of the plasma on the surface of the vessel containing it, such an external current being similar to an external beam of particles discussed above.

Generation of waves by a moving current is a special case of Vavilov-Čerenkov radiation. Morozov has shown that under certain conditions the absorption in plasma of magneto-acoustic waves produced in this way may in principle lead to a very considerable heating of the plasma. Of course the velocity of the current must exceed the velocity of the waves in question. One of the causes of high heating efficiency of a current is the coherence of the waves generated by its different elements. In this respect there exists an analogy between a current and a cluster of charged particles, the radiation of a current being proportional to the square of its strength.

There is another possible way of utilizing the Vavilov-Čerenkov radiation of a current. It is well known that currents excited in plasma, which in virtue of the pinch-effect are usually concentrated in a thin thread, are highly unstable. Therefore in practical applications it is often all-important to sta-

bilize them. If the walls of the vessel containing plasma are conducting, then a displacement of the plasma current towards these walls will induce Foucault currents in them, and these currents will tend to repel the plasma current backwards. Methods of stabilization based on this phenomenon were independently proposed by physicists in different countries and were used in a number of thermonuclear experiments, but have proved to be not very satisfactory. Morozov and Soloviev[14] (1958) have recently proposed to construct the walls of vessels containing plasma not of conducting materials but of such materials, in which velocities of propagation of electromagnetic waves in an appropriate range of frequencies are as small as possible. If a current, flowing in plasma along the surface of such a wall, is displaced towards this surface with a velocity exceeding the velocity of propagation in the wall of waves of a certain frequency, then these waves will be radiated by the current into the wall. The recoil force acting on the current will tend to repel it from the wall and thus to stabilize the current.

I wish to emphasize that I have no definite opinion on possible advantages and disadvantages of methods of heating and of stabilization mentioned or on their technical feasibility. They were selected by me only as examples of possible applications of the general theory, which I have outlined in the beginning. The applications mentioned were necessarily confined to a very limited domain of physics.

I can only hope to have to some extent succeeded to convey to you the impression that there are further possibilities to apply this theory to new and interesting physical problems, and that work done on these lines may be useful in solving these problems or at least getting an insight into the general physical mechanism of some of the relevant phenomena.

1. V. L. Ginzburg and I. Frank, *J. Phys. USSR*, 9 (1945) 353.
2. A. Sommerfeld, *Göttingen Nachr.*, 99 (1904) 363.
3. I. Tamm and I. Frank, *Compt. Rend. Acad. Sci. USSR*, 14 (1937) 109.
4. I. Frank, *J. Phys. USSR*, 7 (1943) 49.
5. V. L. Ginzburg and I. Frank, *Compt. Rend. Acad. Sci. USSR*, 56 (1947) 583.
6. G. Marx and G. Györgyi, *Ann. Physik*, 16 (1955) 241.
7. E. Fermi, *Phys. Rev.*, 57 (1940) 485.
8. D. Pines, *Rev. Mod. Phys.*, 28 (1956) 184.
9. L. Landau, *Zh. Eksperim. i Teor. Fiz.*, 16 (1946) 574.
10. V. L. Ginzburg and V. V. Zhelesniakov, *Astron. Zh.*, 35 (1958) 694.

11. D. Bohm and E. P. Gross, *Phys. Rev.*, 75 (1949) 1864.
12. V. L. Ginzburg, *Compt. Rend. Acad. Sci. USSR*, 56 (1947) 145.
13. A. Morozov, *Plasma Physics and Thermonuclear Reactions*, Moscow, 1958, Vol. 4, p. 331.
14. A. Morozov and L. Soloviev, *Ibid.*, Vol. 4, p. 391.
15. I. Tamm, *J. Phys. USSR*, 1 (1939) 439.

Biography

Igor' Evgen'evič Tamm was born in Vladivostok on July 8, 1895, as the son of Evgenij Tamm, an engineer, and Olga Davydova. He graduated from Moscow State University in 1918, specializing in physics, and immediately commenced an academic career in institutes of higher learning. He was progressively assistant, instructor, lecturer, and professor in charge of chairs, and he has taught in the Crimean and Moscow State Universities, in Polytechnical and Engineering-Physical Institutes, and in the J. M. Sverdlov Communist University. Tamm was awarded the degree of Doctor of Physico-Mathematical Sciences, and he has attained the academic rank of Professor. Since 1934, he has been in charge of the theoretical division of the P. N. Lebedev Institute of Physics of the U.S.S.R. Academy of Sciences.

A decisive influence on his scientific activity was exercised by Prof. L. Mandelstam, under whose guidance he worked a number of years and with whom he was closely associated since 1920, when they met for the first time, and up to the death of Prof. Mandelstam in 1944.

Tamm is an outstanding theoretical physicist, and his early researches were devoted to crystallo-optics and the quantum theory of diffused light in solid bodies. He turned his attention to the theory of relativity and quantum mechanics and he evolved a method for interpreting the interaction of nuclear particles. Together with I. M. Frank, he developed the theoretical interpretation of the radiation of electrons moving through matter faster than the speed of light (the Čerenkov effect), and the theory of showers in cosmic rays. He has also contributed towards methods for the control of thermonuclear reactions. Resulting from his original researches, Tamm has written two important books, *Relativistic Interaction of Elementary Particles* (1935) and *On the Magnetic Moment of the Neutron* (1938).

I. Tamm was elected Corresponding Member of the U.S.S.R. Academy of Sciences in 1933, and in 1953 he became an Academician. He shared the 1946 State Prize with Vavilov, Čerenkov, and Frank, and is a Hero of Socialist Labour. He is also a member of the Polish Academy of Sciences, the American Academy of Arts and Sciences and the Swedish Physical Society.

Physics 1959

EMILIO GINO SEGRÈ

OWEN CHAMBERLAIN

«for their discovery of the antiproton»

Physics 1959

Presentation Speech by Professor E. Hulthén, Chairman of the Nobel Committee for Physics

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

The notion of matter as something built up of very tiny and indivisible atoms is a heritage from classical times. Since, however, experimental research in our days has shown that the atoms in their turn are complicated structures, the notion of indivisibility has been transferred to the so-called elementary particles of which the atom is composed, in the hope of therewith having reached the ultimate limit for the division of matter.

However, the different kinds of elementary particles showed an alarming tendency to increase in number – something which is at variance with the attractive idea that matter is built up of one or at most two kinds of particles.

Among the most successful and noteworthy attempts to interpret this situation is Dirac's theory of particles and antiparticles, which may be designated, almost, as each other's mirror images. Both kinds of particles are conceived as arising through the formation of pairs and as reciprocally annihilating each other. The world in which we find ourselves belongs, by chance, to the one kind of particles, among which sporadically occurring antiparticles are very quickly destroyed. On account of the mirror symmetry it would be very difficult to decide whether a remote star or galaxy belonged to the one or the other kind of matter.

There were probably very few physicists who at first ascribed to this side of Dirac's otherwise very valuable theory any real import until, quite suddenly and unexpectedly, the first antiparticle, the positive electron, was discovered by Anderson in cosmic radiation in the year 1931. Continued investigations showed that the new particle behaved in all respects according to Dirac's theory: that it was manifested, namely, in connection with some energetic process, always together with an ordinary electron, and that it disappeared in the same way and with equal suddenness. Today nothing is better known and clearly elucidated than this process of pair formation and annihilation.

Nonetheless it now seemed desirable to test the validity of this remarkable

theory upon the antiparticle to the proton, the nucleus of the hydrogen atom – a process which required, however, about 2,000 times as great an amount of energy. Such quantas of energy do, certainly, occur in cosmic radiation, but in such a random way that it was finally realized that the only systematic way of investigating the process was through the controlled production of the antiproton by means of an accelerator with a sufficiently high capacity.

It has been said of the *Bevatron*, the great proton accelerator at Berkeley University in California, that it was constructed chiefly with a view to the production of antiprotons. This is perhaps an exaggeration, but in so far correct as its peak achievement, 6 milliard electron volts, was set with a view to the energy required for the pair formation of protons–antiprotons. That it was constructed in Berkeley was due to the tradition established there ever since Lawrence built the first cyclotron and McMillan developed the principle for the synchronization of relativistic particles.

But even if antiproton research was thus first made possible through this technologically very impressive machine, the actual discovery and investigation of the antiproton was chiefly the merit of Chamberlain and Segrè. With similar methods an antiparticle to the neutron has subsequently been discovered, a discovery whose importance lies in the fact that the concept of the antiparticle was thereby extended to include also the neutral elementary particles.

Professor Emilio Segrè, Professor Owen Chamberlain. Your discovery of the antiproton was made possible through the excellent resources at the Radiation Laboratory in Berkeley. It is, however, your ingenious methods for the detection and analysis of the new particle that the Royal Swedish Academy of Sciences wishes to recognize on this occasion.

I need surely not remind you, Professor Segrè, of the occasion, twenty-one years ago, when your compatriot Enrico Fermi received his Nobel Prize in this selfsame place. You and he were intimate friends and you had been collaborating with great success. Both of you belonged to that group of distinguished Italian scientists that was westward bound in those days.

Also you, Professor Chamberlain, must surely have an intimate and abiding recollection of your years together with Fermi in Chicago.

Gentlemen, I now ask you to receive your prize from the hands of His Majesty the King.

OWEN CHAMBERLAIN

The early antiproton work

Nobel Lecture, December 11, 1959

It is a pleasure to have this opportunity of describing to you the early work on the antiproton, particularly that part which led to our first paper on the subject, in 1955, *Observation of Antiprotons*, by Professor Emilio Segrè, Dr. Clyde Wiegand, Professor Thomas Ypsilantis, and myself[1]. At the same time, I would like to convey my impressions of the conclusions that can be drawn purely from the *existence* of the antiproton. The later experiments, that have led to results quite unsuspected in the beginning, will be discussed in Professor Segrè's Nobel Lecture.

The first inkling of the idea that each electrically charged particle might have an oppositely charged counterpart may be seen in the theory of P.A.M. Dirac[2], proposed in 1930. Perhaps you will allow me to oversimplify the situation, and to refer to the relativistic equation relating momentum p to energy E, namely,

$$E^2 = p^2c^2 + m^2c^4$$

where p = momentum, E = energy, m = mass, c = velocity of light.

It is clear that this equation, as it stands, permits either negative or positive values for the energy E. In everyday life we do not see particles of negative energy. Nevertheless the theory of Dirac encompassed electrons of negative energy. Dirac observed that an electron, which is electrically negative, if in a negative-energy state, would have some of the properties of a positively charged particle. In fact, at one time Dirac believed that the states of negative energy might be somehow related to protons, which are electrically positive, and not at all to electrons. As the Dirac equation was further studied, it became apparent that the negative energy-states represented a new phenomenon, one rather difficult to understand.

In Dirac's theory the negative energy-states were assumed to exist, but were taken to be completely filled with electrons normally. The Pauli exclusion principle (which forbids that two electrons should occupy the same state) was used to explain why an electron in a positive-energy state would

not normally drop into a filled state of negative energy. Perhaps this situation may be made clearer by reference to the following diagrams (see Fig. 1):

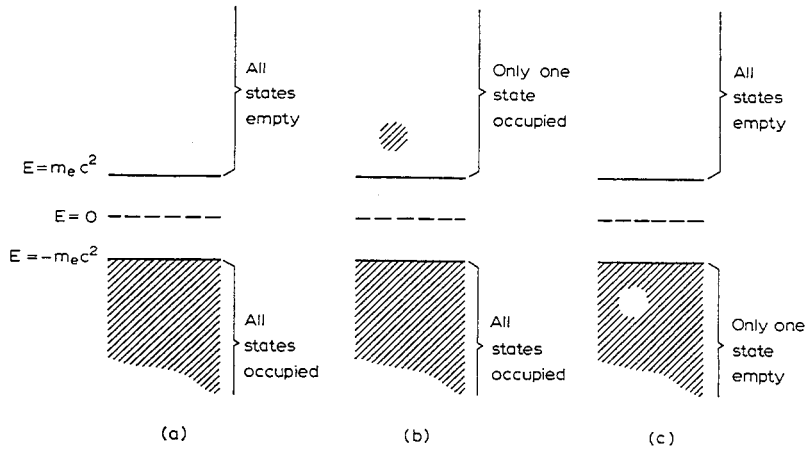


Fig. 1. Diagram of energy levels in the Dirac theory. Part (*a*) represents a vacuum, (*b*) a state with one electron present, and (*c*) a state with one positron.

The sketch on the left shows an empty vacuum, with all the positive-energy states empty and all the negative-energy states filled. Notice that the states of highest energy are represented as highest in the diagram. The second view shows a state with only one electron present, the filled negative-energy states being unobservable. The last view shows a state in which no positive-energy electron is present, but one electron is missing from the states of negative energy. Experimentally, this would correspond to one positive charge being present. This positive charge was shown to move exactly as if its mass were the same as that of the electron.

The Dirac theory was taken much more seriously after Professor C. D. Anderson's discovery of the positron[3] in 1932. This was the first of the antiparticles to be found, for the positron was really the antielectron, and might have been so named. The positron has the same mass as the electron, but has an opposite charge. Electron and positron are sometimes born as a pair, as in the process we call pair production by gamma rays. Electron and positron die as a pair, in a process we call annihilation. The electron and positron eat each other up, so to speak; what remains is energy in the form of gamma rays.

The Dirac theory was extremely successful in explaining electron and pos-

itron phenomena. Therefore, the question naturally arose: cannot the proton also be treated according to the Dirac theory? The proton, like the electron, has spin of one-half, and obeys the Pauli principle, i.e. two or more protons are never found simultaneously in exactly the same state. Still, the proton in some respects did not fit the Dirac pattern as well as the electron did. For example, the magnetic moment of the proton was not in accord with the theoretical value, while that of the electron was extremely close to it. While Professor Pauli did introduce a form of the Dirac theory that allowed agreement with the observed magnetic moment of the proton, still the theory had lost some of its simplicity.

If the Dirac theory were correct for describing protons as well as electrons, then it clearly would call for the existence of antiprotons as well as of protons, just as it had demanded the existence of positrons as well as of electrons.

According to the Dirac theory, the antiproton should have the properties indicated in Table 1.

Table 1. Anticipated properties of the antiproton.

Property	*Values expected*
Mass	Same as the proton mass
Electric charge	Negative, opposite to the proton, equal to the electron charge
Spin	1/2 unit, same as the proton
Magnetic moment	Equal in magnitude, opposite in direction to the proton
Statistics	Fermi statistics, like the proton
Intrinsic parity	Opposite to the proton

Here, I would like to make a small digression. Physicists had asked themselves why neutrons and protons may turn into one another but are never completely destroyed. For example, why cannot a reaction such as the one shown here be observed?

$$\text{proton} \not\rightarrow \text{positron} + 2 \text{ gamma rays}$$

Although the answer in deeper terms was not forthcoming, it was generally believed that there must be a fundamental rule specifically providing that the heavy particles are never destroyed (or created) singly. There remained the possibility, however, that one might make a pair, proton and antiproton, at the same time. Similarly, proton and antiproton might annihilate each other,

thus dying in pairs, one particle (the proton) and one antiparticle (the antiproton).

Thus it was expected that the creation of an antiproton would necessitate the simultaneous creation of a proton or a neutron. This meant that the energy necessary to produce a proton–antiproton pair would be twice the proton rest energy, or about 2 billion electron volts. It was also known that high-energy collisions were necessary, such as those between high-energy particles and particles at rest. A likely process might be one in which a very fast proton struck a stationary neutron. The proton would have to have about 6 billion electron volts of energy in order that 2 billion electron volts might be available to make new particles.

For many years physicists working with cosmic rays kept a sharp watch for antiprotons, for it was known that there were particles in the cosmic radiation with sufficient energy to produce antiprotons. Some events were observed[4] which might have involved an antiproton, but it was not possible to determine with certainty the presence of an antiproton. I should also mention that the possible antiproton events were remarkably few. As the years passed, people began to have serious doubts that the antiproton could indeed be produced.

I am told that the men responsible for the early design of the Bevatron, the large accelerator in Berkeley, California, were well aware that about 6 billion electron volts were needed to produce antiprotons, and that this fact was taken into consideration when the Bevatron energy, just 6 billion electron volts, was decided upon. Professor Edwin McMillan and the late Professor Ernest Lawrence were among the men who selected the energy of the Bevatron and were responsible for initiating its construction.

During the years when the Bevatron was under construction, doubts as to the existence of the antiproton tended to increase the interest in an experimental demonstration of the antiproton. To make this demonstration, it was necessary to show that after the Bevatron beam struck a piece of material, there were present among the particles issuing from nuclear collisions some particles of the same mass as that of the proton, but electrically negative, i.e. with an electric charge opposite to that of the proton. It is not difficult to sort out particles that are electrically negative, for in a magnetic field the negative particles are deflected in one direction, while positive particles are deflected in the opposite direction.

From the start it was understood that at least two independent quantities must be measured for the same particle, if the mass of the particle was to be

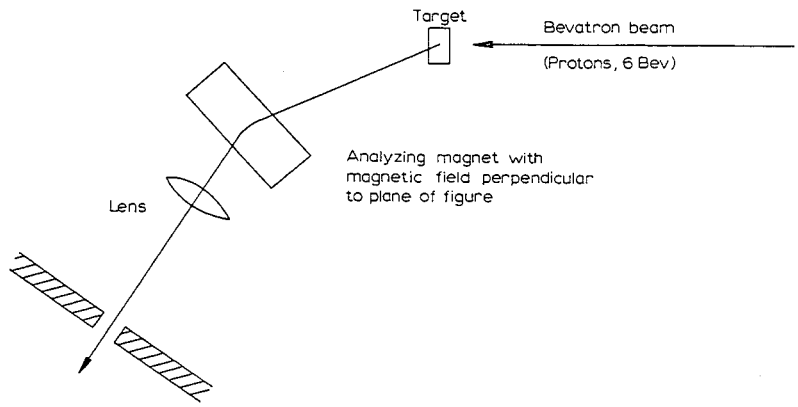


Fig. 2. Simple magnetic spectrograph for producing a beam of charged particles with known momentum.

determined. Any two of the following quantities would be adequate: momentum, energy, velocity, range (distance of penetration into a substance before coming to rest). For example, to measure only the momentum would be insufficient, because particles with various masses may have the same momentum. However, if two quantities are measured, then the mass of the particle can be calculated. After several possibilities were considered, it was decided that the two quantities to be measured should be momentum and velocity.

Perhaps I may remind you of the nonrelativistic relationship between momentum and velocity, namely

$$p = mv$$

where p = momentum, m = mass, v = velocity. Clearly, we may obtain the mass of a particle by measuring momentum and velocity, obtaining the mass as $m = p/v$. For particles travelling at nearly the speed of light, the expressions are not as simple as these; but it is still true that if momentum and velocity are simultaneously measured, then the mass of the particle may be determined. The relationships are

$$p = \frac{mv}{\sqrt{1 - v^2/c^2}} \qquad m = \frac{p\sqrt{1 - v^2/c^2}}{v}$$

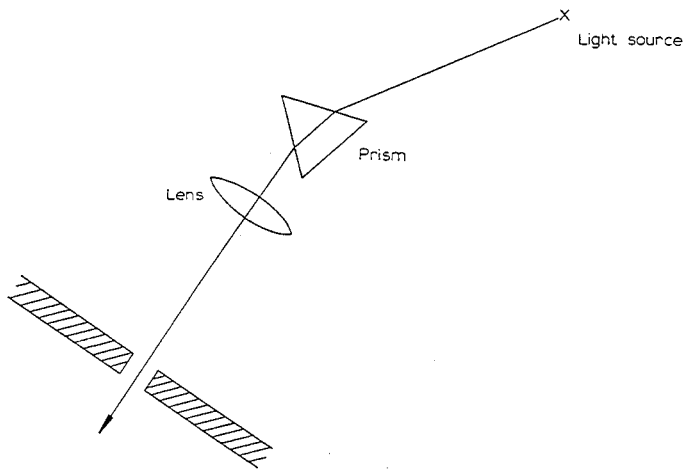


Fig. 3. Optical spectrograph.

The well-established method of measuring the momentum of atomic particles is to measure their deflection in a known magnetic field. In fact, one may set up a system of magnets through which only particles of a certain momentum interval pass. Such a system may be arranged as shown in the diagram of Fig. 2.

The high-energy proton beam strikes a target, which might be a piece of copper. Fragments of nuclear collisions emerge from the target in all directions, some of which enter the region of the magnetic field, the rest being wasted. For certain values of momentum, negative particles are deflected by the correct amount and pass through the aperture in the collimator. Other particles are stopped by the walls of the collimator.

A lens is also shown in the same figure. The use of magnetic lenses was suggested to us by Dr. Oreste Piccioni, who pointed out that a greater number of useful particles would pass through the system if lenses were employed.

I will not describe the lenses in detail, except to say that each lens is composed of three quadrupole magnets. The lens action is very similar to the action of a simple optical lens for light. In fact, the system of magnets we are discussing acts, for charged particles, just as the system shown in Fig. 3 acts for light.

Here a light source replaces the target, and the analyzing magnet is replaced by a glass prism. Through the final slit, the system shown would only

allow light to pass in a narrow range of wavelengths or colors. I will refer to this system as a spectrograph, whether it be for light or for charged particles.

I have said that only negative particles would be present in the beam formed by the magnetic spectrograph, since positive particles would be deflected in the opposite direction and would not pass through the collimator. By reversing the magnetic field, however, we may use the same spectrograph to select positive particles, if these are desired.

The final arrangement of magnets that we have used consists of two such spectrographs, one following the other. It is shown in Fig. 4.

The first spectrograph extends from the target to the first focus point. The second spectrograph extends from the first to the second focus point. The first spectrograph acted as a preliminary sorting device to provide a relatively pure beam of particles for the second spectrograph, where the experiment itself was performed.

Now let us discuss the velocity of the antiprotons that may be found. The most copious production of antiprotons was expected to be at a velocity of 87 percent of the velocity of light, while the other background particles, such as mesons, would be travelling at approximately 99 percent of the velocity of light. To make the experiment somewhat easier, slightly slower antiprotons were chosen, with $v = 0.78\ c$, or 78 percent of the velocity of light.

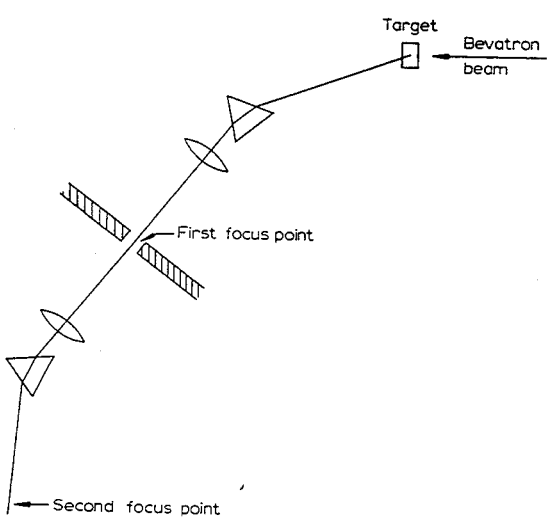


Fig. 4. Final arrangement of magnets, represented in optical analogy.

One of the simplest means of determining velocity is to measure the time of flight between two counters, and this was the method used. A counter was placed at the first focus point, and another counter at the second focus point. The spacing of these counters was 12 meters. Mesons, which travel near the speed of light, would take 40 millimicroseconds (40 billionths of a second) to pass over this distance, but antiprotons of the same momentum would take 51 millimicroseconds. The difference of 11 millimicroseconds is not great, but it is within the capability of modern electronic circuits to distinguish with certainty between the two time intervals.

Unfortunately, when many mesons are passing through the system in close succession, it will sometimes happen that two mesons have exactly the right spacing between them to imitate an antiproton, i.e. just 11 millimicroseconds. We will call this the accidental coincidence of two mesons. To avoid having the electronic circuits «fooled» by two mesons in close succession, it was considered important to have another method of determining velocity in use at the same time. For this purpose, a Čerenkov counter was constructed. In fact, the study of the Čerenkov counter was one of the first steps taken in the preparation of the experiment. I would like to describe it here.

The basic principles of the Čerenkov radiation are familiar, since Professor Čerenkov must have been speaking from this rostrum just one year ago, but those principles especially relevant to the Čerenkov counter used in the antiproton experiment should be reviewed here. In an optical medium such as

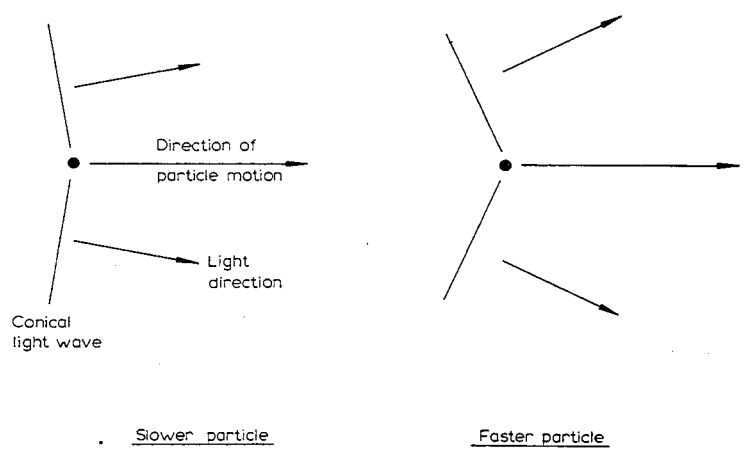


Fig. 5. Diagram of Čerenkov radiation. The angle of emission of Čerenkov light depends on the speed of the charged particle.

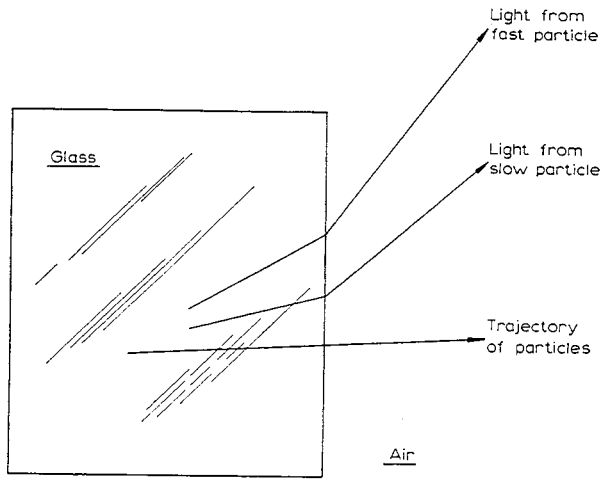


Fig. 6. Refraction of Čerenkov radiation at the interface between glass and air.

glass, light moves more slowly than in a vacuum. Although no particle may move faster than the speed of light in vacuum, a particle may move faster than the speed of light in glass. When this happens with an electrically charged particle, Čerenkov light is emitted, at different angles for particles of different velocity. This is illustrated in Fig. 5.

When the Čerenkov light passes out of the piece of glass into air, it is refracted at the surface, but it continues to be true that the direction of the light in air is dependent upon the velocity of the particle, as is shown in Fig. 6.

A cylindrical mirror was arranged to reflect only light emitted in the right direction. This light was then refocused within a small region where it could be detected by a photomultiplier. The design of this counter is shown in Fig. 7.

For the purpose of understanding the action of the counter, we may think of one photomultiplier, placed along the axis of the instrument, catching the Čerenkov radiation emitted by particles travelling parallel to the axis with the correct velocity. In practice, the light was divided into three parts by three flat mirrors, to avoid having the beam of particles traverse the photomultiplier. In the first experiment the material of the Čerenkov radiator was fused quartz.

At the time that I first designed one of these « velocity selecting » counters I thought that the design was original with me. I later realized, however, that the design is the same as that of Dr. Samuel Lindenbaum. I had heard

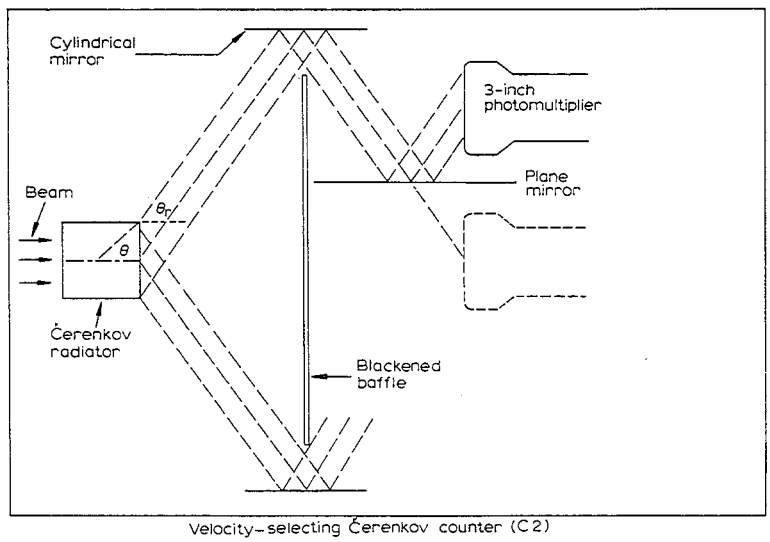


Fig. 7. View of the velocity-selecting Čerenkov counter.

discussion of the Lindenbaum counter, but had not then understood it.

We assumed that our greatest difficulty in identifying antiprotons with certainty would be the possibility of a mistake in the determination of velocity. Accordingly, we also used another Čerenkov counter, of a more conventional design. It was a container of liquid, with an index of refraction of 1.28, viewed directly by a photomultiplier. Particles moving as slowly as our hoped-for antiprotons would produce no Čerenkov light in this material, since they would not be moving faster than light in that medium. However, faster particles, such as mesons and electrons, would produce Čerenkov light in this material and would give rise to a pulse at the output of this counter. Since this counter was to serve as a warning against particles faster than antiprotons, we may call it a guard counter. The liquid used in the guard counter was a fluorocarbon.

As the whole system was assembled, the plan of the apparatus looked as follows (see Fig. 8):

Our electronic circuits gave out an electrical signal each time the following events occurred within a short time interval: the two counters indicated as S1 and S2 counted with a relative time delay of 51 ± 5 millimicroseconds, and the velocity-selecting counter C2 also gave a signal in at least two of its three photomultipliers. Whenever these events occurred, it meant a potential

antiproton had been seen in the apparatus. This signal was used to start an oscilloscope, from which a picture of the pulses in the counters S1, S2, and the guard counter C1 could be obtained.

Three such pictures are shown in Fig. 9. The horizontal motion of the dot on the oscilloscope screen is uniform, so the curve shown represents an electrical signal as a function of time. The first pulse on each curve (oscilloscope trace) is due to the counter S1, the second is due to S2, and the third (if present) to the guard counter C1.

The pulses themselves look rather broad in this figure, but by careful measurement of the position of the pulse at its beginning, the difference in time between pulses can be measured to an accuracy of about one millimicrosecond.

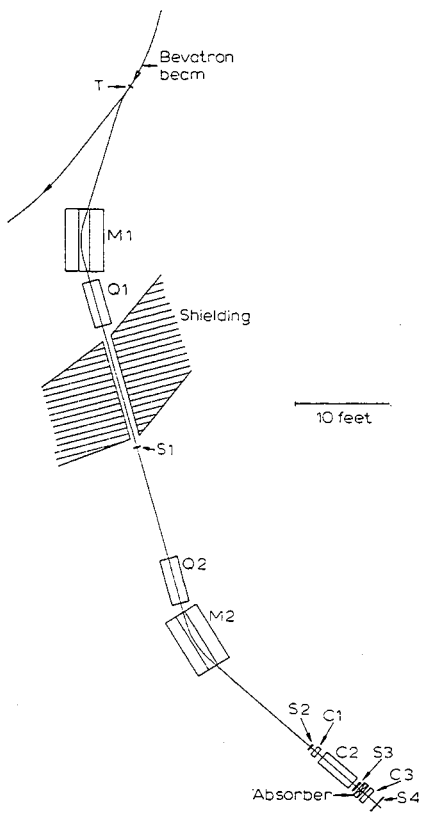


Fig. 8. Plan view of the apparatus used in the discovery of antiprotons. T indicates the target, Q1 and Q2 are magnetic lenses, C1 is the guard counter, and C2 is the velocity-selecting Čerenkov counter. The last two counters (C3 and S4) may be ignored.

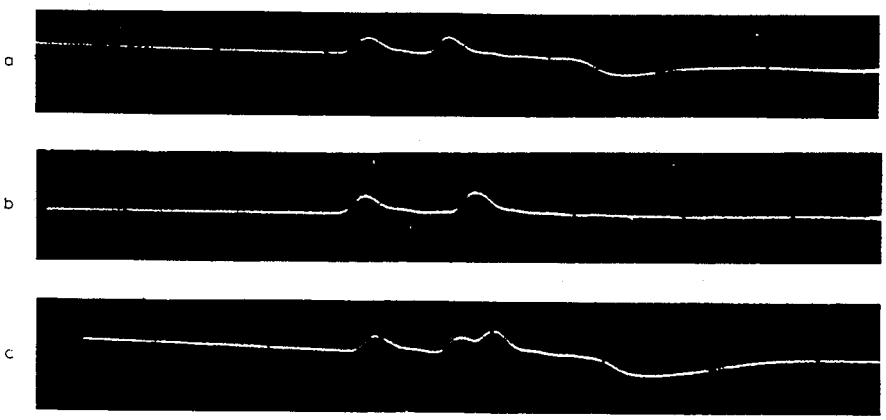


Fig. 9. Oscilloscope traces produced by (*a*) a meson, (*b*) an antiproton, (*c*) an accidental coincidence. The first two pulses are due to scintillation counters S1 and S2. The separation is noticeably different for a meson and for an antiproton.

Since the momentum is accurately known from the magnet arrangement, and since each time measurement determines the velocity of the particle in question, we have in effect a rough mass measurement of each antiproton.

Fig. 10 (*a*) shows the distribution of the differences in time for mesons. It was obtained when the system was adjusted to register mesons. Fig. 10 (*b*), the distribution of times of flight for antiprotons, was obtained when the system was adjusted in the normal way, to accept antiprotons but reject mesons. The distribution is well-peaked at the expected time of flight of antiprotons. Fig. 10 (*c*) shows a representative group of accidental coincidences.

In the early stages of the experiment, we were in some danger of failing to observe antiprotons, because we had set our apparatus for a slightly different mass than the proton mass, without our intending to do so. It was essentially an error in the alignment of the apparatus. Fortunately, we were able to reverse the magnetic fields in all the magnets and thus look for protons. Discovering very few, we soon found the error of alignment. After suitable adjustments, protons were easily observed and, when the magnetic fields were again reversed, possible antiproton events appeared.

Following approximately twenty such possible events, the equipment was subjected to a series of tests designed to determine whether the antiprotons seen were completely real, or whether they might be due to some failure of the apparatus. One effective test involved our intentionally tuning the sys-

tem to a somewhat different mass than the proton mass to test whether it was true that the antiprotons no longer registered. This indicated that the apparatus really was sensitive to the mass of the particles. Another important test entailed lowering the energy of the beam of the Bevatron, to make sure that the supposed antiprotons disappeared when the beam energy was sufficiently low.

Fig. 11 shows the fraction of antiprotons in the beam as a function of the energy of the original proton beam. Later calculations showed that this result was reasonable, and that a few antiprotons should be produced at 5 BeV, as was observed. These tests convinced us that we had indeed observed the production of antiprotons.

The methods I have described are now well-known to physicists working in this field. One may ask, why is the present work considered such an accomplishment? The answer lies partly in the fact that the antiprotons are

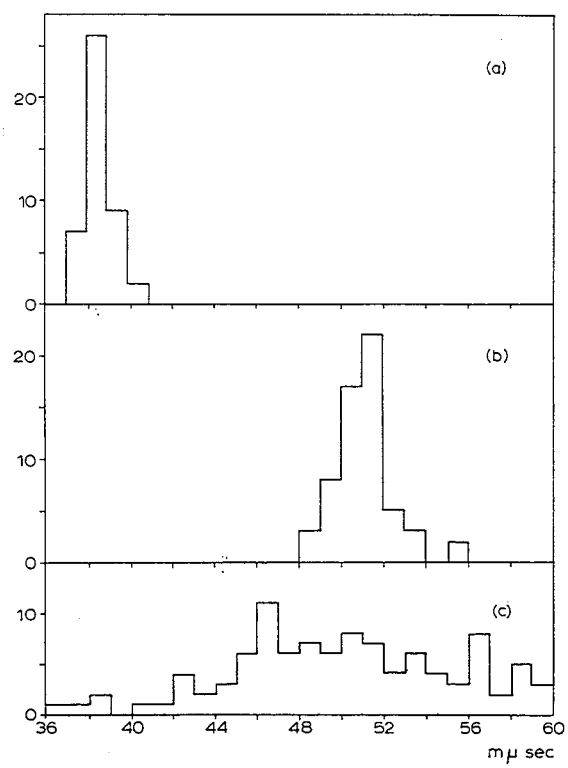


Fig. 10. (*a*) Histogram of times of flight for mesons; (*b*) histogram of times of flight for antiprotons; (*c*) apparent flight times for accidental coincidences.

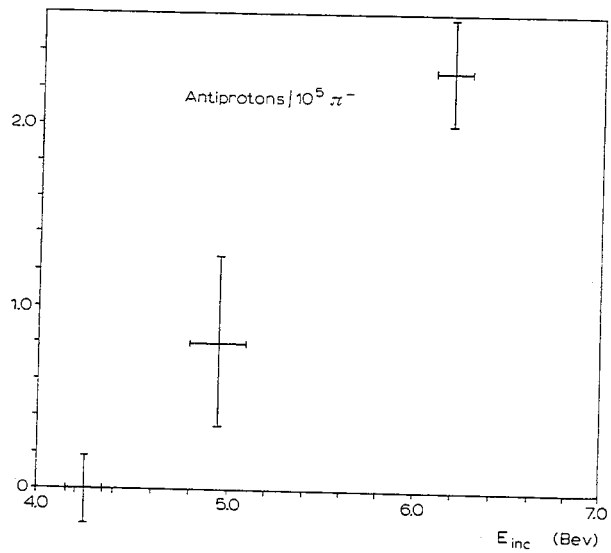


Fig. 11. Excitation curve for the production of antiprotons, relative to meson production, as a function of the energy of the protons in the Bevatron.

extremely rare in a beam of many other particles. Early rough estimates indicated that an antiproton might be produced only once in each million collisions within the Bevatron target. At the same time, it was estimated that antiprotons would largely go in a forward direction with respect to the beam, while to a great extent other particles would go in all directions. This effect was estimated as giving a ten-fold enrichment in the forward direction, meaning that there would be conceivably one antiproton among each 100,000 of the other particles, which were mostly mesons. Fortunately the antiprotons were slightly more numerous, being 1 in 30,000 particles in the magnetically analyzed beam. If there had been appreciably fewer antiprotons, we might altogether have missed seeing them on the first try. As it was, we saw only one antiproton every fifteen minutes with the first apparatus.

In conjunction with the counter experiment I have described, we also collaborated in a related experiment[5] using photographic emulsions. Within a short time after the first counter experiments were completed, examples of the annihilation phenomenon were seen, in which the antiproton and a proton or neutron from an ordinary nucleus die simultaneously. In their place emerge about five π mesons.

Fig. 12 shows how one of the annihilation events looked under the microscope in the photographic emulsion.

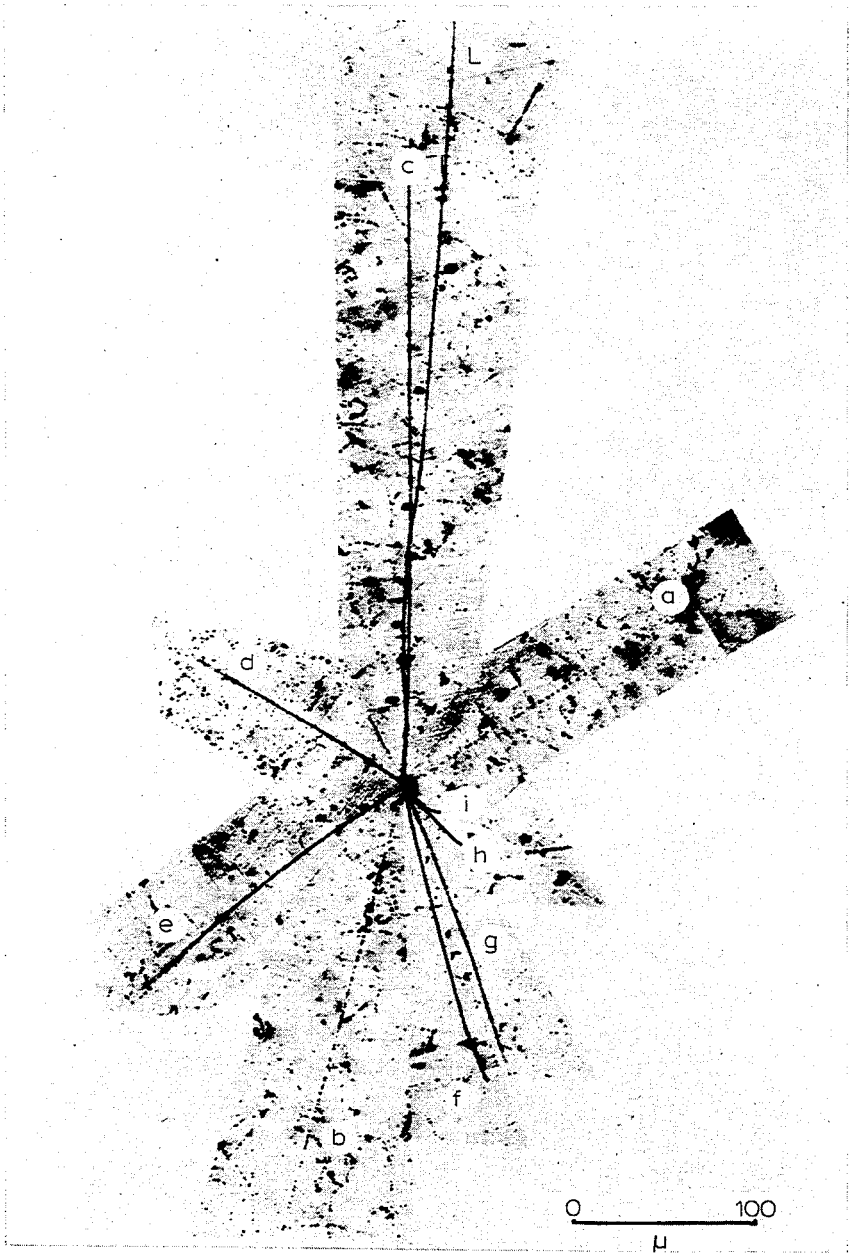


Fig. 12. Reproduction of an antiproton star as seen in nuclear emulsion. L is the incoming antiproton track. Tracks a and b are due to mesons. The remaining tracks are due to protons or alpha particles.

The work with photographic emulsions involved many people, among whom I would like to mention especially Professor Gerson Goldhaber and Professor Edoardo Amaldi from Rome. I would like to mention that it was a young Swedish physicist, Dr. A. G. Ekspong, now at the University of Uppsala but then working with us in Berkeley, who observed under the microscope the first of the annihilation events with very large visible release of energy. That star gave the final visual proof through the phenomenon of annihilation that we were dealing with antiprotons, rather than with any other similar particle. The large release of energy showed not only that the incoming particle (antiproton) died in the production of the star, but that additional energy must have been supplied by the death of an ordinary nucleon (neutron or proton).

By the conclusion of these early experiments, the existence of the antiproton had been established beyond doubt; its charge was of course negative; its mass was established as equal to the proton mass, to within 3 percent. Furthermore, the data on production in the Bevatron at different proton-beam energies clearly indicated that antiprotons are born in pairs with ordinary protons or neutrons. The emulsions showed that they die in the annihilation process.

Since the proton and neutron are close sisters, it was expected that the discovery of the antineutron would quickly follow that of the antiproton. In fact, it is natural to infer that antiparticles of all charged particles exist. This inference also extends to the neutral particles that have some non-zero quantum number analogous to electric charge. Today there is much experimental evidence, as well as strong theoretical suggestions, to confirm this.

In conclusion, I would like to name some of the many people besides Professor Segrè and myself who have been responsible for the success of this work. Obviously, our co-authors, Dr. Clyde Wiegand and Professor Thomas Ypsilantis, have contributed heavily. I wish to mention particularly the work of Dr. Wiegand, for his part was absolutely essential to our achievement. For the very existence of the large accelerator, we owe a great debt to the late Professor Ernest O. Lawrence, to Professor Edwin M. McMillan, now Director of the Lawrence Radiation Laboratory, and especially to Dr. Edward Lofgren, who has been responsible for the construction and operation of the Bevatron. I have mentioned Dr. Oreste Piccioni's contribution with respect to the use of quadrupole lenses. I should also cite the important work of Dr. Herbert M. Steiner, who collaborated with us throughout the whole experiment. Needless to say, we received wonderful cooperation

from the men who operate the Bevatron and from the engineers and technicians who aided us in many ways with the design, the construction, and the installation of the apparatus. Our own work was built directly upon the previous accomplishments of many eminent scientists, including especially P. A. M. Dirac, C. D. Anderson, E. M. McMillan, and P. A. Čerenkov.

Other important work closely related to the same subject has occupied Professor Amaldi with his colleagues at the University of Rome, and Professor Gerson Goldhaber with his collaborators in Berkeley. I think it is clear, therefore, that the work of many people is being honored today.

1. O. Chamberlain, E. Segrè, C. E. Wiegand, and T. Ypsilantis, *Phys. Rev.*, 100 (1955) 947.
2. P. A. M. Dirac, *Proc. Roy. Soc. London*, A 126 (1930) 360; A 133 (1931) 60.
3. C. D. Anderson, *Science*, 76 (1932) 238.
4. E. Hayward, *Phys. Rev.*, 72 (1947) 937; H. S. Bridge, H. Courant, H. DeStaebler, and B. Rossi, *Phys. Rev.*, 95 (1954) 1101; E. Amaldi, C. Castagnoli, G. Cortini, C. Franzinetti, and A. Manfredini, *Nuovo Cimento*, 1 (1955) 492.
5. O. Chamberlain. W. W. Chupp, G. Goldhaber, E. Segrè, C. E. Wiegand, E. Amaldi, G. Baroni, C. Castagnoli, C. Franzinetti, and A. Manfredini, *Phys. Rev.*, 101 (1956) 909; O. Chamberlain, W. W. Chupp, A. G. Ekspong, G. Goldhaber, S. Goldhaber, E. J. Lofgren, E. Segrè, C. E. Wiegand, E. Amaldi, G. Baroni, C. Castagnoli, C. Franzinetti, and A. Manfredini, *Phys. Rev.*, 102 (1956) 921.

Biography

Owen Chamberlain was born in San Francisco on July 10, 1920. His father was W. Edward Chamberlain, a prominent radiologist with an interest in physics. His mother's maiden name was Genevieve Lucinda Owen.

He obtained his bachelor's degree at Dartmouth College in 1941. He entered graduate school in physics at the University of California, but his studies were interrupted by the involvement of the United States in World War II. In early 1942 he joined the Manhattan Project, the U.S. Government organization for the construction of the atomic bomb. Within the Manhattan Project he worked under Professor Emilio Segrè, both in Berkeley, California, and in Los Alamos, New Mexico, investigating nuclear cross sections for intermediate-energy neutrons and the spontaneous fission of heavy elements. In 1946 he resumed graduate work at the University of Chicago where, under the inspired guidance of the late Professor Enrico Fermi, he worked toward his doctorate. He completed experimental work on the diffraction of slow neutrons in liquids in 1948 and his doctor's degree was awarded in 1949 by the University of Chicago.

In 1948 he accepted a teaching position at the University of California in Berkeley. His research work includes extensive studies of proton–proton scattering, undertaken with Professor Segrè and Dr. Clyde Wiegand, and an important series of experiments on polarization effects in proton scattering, culminating in the triple-scattering experiments with Professor Segrè, Dr. Wiegand, Dr. Thomas Ypsilantis, and Dr. Robert D. Tripp. In 1955 he participated with Dr. Wiegand, Professor Segrè, and Dr. Ypsilantis in the discovery of the antiproton. Since that time he has taken part in a number of experiments designed to determine the interactions of antiprotons with hydrogen and deuterium, the production of antineutrons from antiprotons, and the scattering of π mesons.

He is a Fellow of the American Physical Society and a member of the National Academy of Sciences. He was awarded a Guggenheim Fellowship in 1957 for the purpose of doing studies in the physics of antinucleons at the University of Rome. He was appointed Professor of Physics at the Univer-

sity of California, Berkeley, in 1958, and served as Loeb Lecturer at Harvard University in 1959.

In 1943 he married Beatrice Babette Copper. He has three daughters and one son.

EMILIO G. SEGRÈ

Properties of antinucleons

Nobel Lecture, December 11, 1959

I must begin by thanking the Swedish Academy for the great honor they have bestowed on me. The names of the previous recipients of the Nobel Award, while lending great prestige to the Award, make me feel humble and dubious about my merits to join the company. However, I can only repeat my gratitude and think that my constant devotion to science may have something to do with the choice, apart from any success, in which there is perforce an element of luck. At the onset I must also mention the names of two people who have had, in different ways, a very great influence upon all my work. Of Enrico Fermi I would only say, quoting Dante as he himself might have done,

Tu se' lo mio maestro e il mio autore;
Tu se' solo colui da cui io tolsi
Lo bello stilo che mi ha fatto onore.

Thou art my master and my author;
thou alone art he from whom I took
the good style that hath done me honor.

I learned from him not only a good part of the physics I know, but above all an attitude towards science which has affected all my work. Ernest Orlando Lawrence created the instruments with which most of my work was done. Although I belong scientifically to a different tradition and outlook, it was only through the instruments developed at his instigation and under his leadership that most of my own researches became possible. This is especially true for the last one: the antiproton.

By 1954 the Bevatron had been developed and tested. It had been purposely planned for an energy above the threshold for forming nucleon–antinucleon pairs, and many physicists including my colleagues and I naturally thought of means for hunting the elusive antiproton. Although its existence was very probable, a definite experimental proof was lacking and, being

aware of the crucial importance of the problem for the extension of Dirac's theory from the electron to the nucleon, we tried to design an experiment which would give a definite answer[1]. The final apparatus has been described in the preceding lecture by Dr. Chamberlain[2].

Other experiments involving photographic detection were also planned at that time and came to fruition soon after the success of the first experiment[3].

Dr. Chamberlain has described to you what an antiproton is and how it was found, and I have nothing to add to his lecture on these matters.

The properties used for the identification of the antiproton were predicted by Dirac long ago and were used as a guide in finding the particle. However, once it was found we faced a host of new problems, and it is to those that I will direct the rest of my speech.

I will be very brief concerning the experimental developments.

Here great emphasis has been put on the development of better antiprotons beams. By « better » I mean beams in which there are more antiprotons per unit time and in which the ratio of the number of antiprotons to unwanted particles is higher. It suffices to say that now it is possible to have at Berkeley, beams with about 10 antiprotons per minute instead of one every fifteen minutes as in 1955 and beams in which antiprotons are about one in ten particles instead of one in 50,000 as in 1955. The improved beams allow more difficult and complicated experiments, and the development of electronics and bubble chambers has kept pace with the increased possibilities. I may add that the complications in which we are entering now are by no means a cause of joy to the experimenters who have to cope with them, and that they are properly considered as the heavy price to be paid in order to obtain more detailed physical information.

Some of the problems raised by the identification of the antiproton have a predictable solution, although the prediction does not derive from anything as solid as Dirac's theory. We could, for instance, expect with complete confidence the existence of the antineutron and of all the antiparticles of the baryons, although it might require considerable skill to find them. In fact, antineutrons are certainly formed copiously at the Bevatron but the primary antineutrons are very difficult to identify. For this reason immediately after the discovery of the antiproton, it was suggested that the antineutron should be found by investigating the charge exchange reaction in which a proton and an antiproton give a neutron and an antineutron[4]. In a very ingenious and elegant counter experiment, Cork, Lambertson, Piccioni, and Wenzel

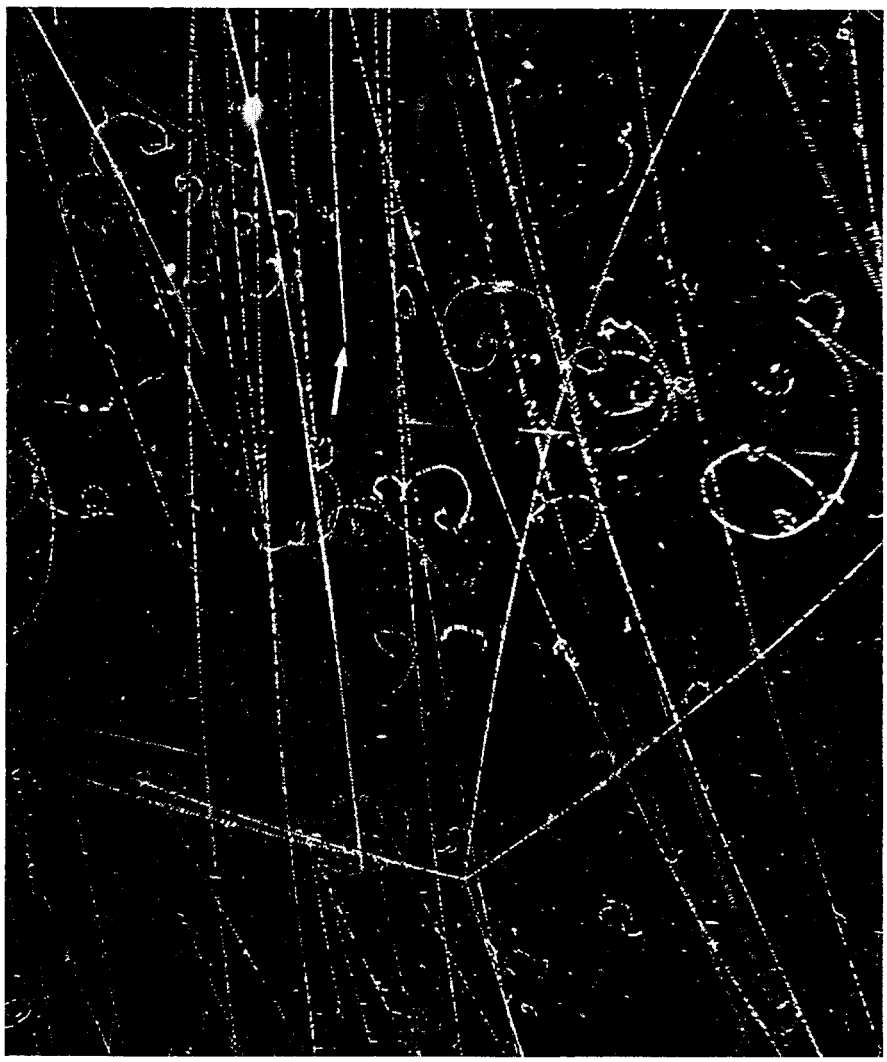


Fig. 1. An antiproton enters a propane bubble chamber, and at the point marked with an arrow undergoes charge exchange. The antineutron originates the annihilation star (directly below). Density of propane 0.42 g/cm^3. Real distance between charge exchange and origin of star 9.5 cm. $T_{\bar{p}}$ at charge exchange $\sim$ 50 MeV. (From Agnew *et al.*[6])

did demonstrate the existence of the antineutron some time ago[5]. Their method was based on a counter technique and uses the reaction

$$p + \bar{p} \rightarrow n + \bar{n}$$

which is called charge exchange because we can interpret it as the passage of the electric charge from the proton to the antiproton. The product antineutron is recognizable by its annihilation properties. Namely, an antineutron on annihilation forms an annihilation star extremely similar to an antiproton star. Instead of reproducing their experimental arrangement, I will show in a slide (Fig. 1) a graphical picture of these phenomena as observed in a bubble chamber by the joint efforts of Professor Wilson Powell and his group, and my own group[6].

Similarly, the antilambda was found by Baldo-Ceolin and Prowse[7] in photographic emulsions exposed to a pion beam and was confirmed in the hydrogen bubble chamber. Also the antisigma has been recently seen in a hydrogen bubble chamber by the Alvarez group in Berkeley[8].

It is also possible to predict with certainty some of the nucleonic properties of the antinucleons, specifically the spin, I-spin, 3rd component of the I-spin, and parity to be those shown in Table 1.

Table 1. Spin, parity, I-spin of nucleons, and antinucleons.

	Proton	*Neutron*	*Anti-proton*	*Anti-neutron*
Spin, S	1/2	1/2	1/2	1/2
I-spin, T	1/2	1/2	1/2	1/2
Third component of I-spin, T_3	1/2	−1/2	−1/2	1/2
Parity	+	+	−	−

But in addition to these interesting questions of systematics of particles, which can be summarized by the diagram shown in Fig. 2, there are problems for which we know much less what to expect because they involve more than general symmetry properties. They require a fairly detailed knowledge of interactions and subnuclear structure which at present we do not have. Indeed these are the most interesting and challenging problems.

For instance, we know that a nucleon and an antinucleon may annihilate each other, but what are the products of the annihilation? What is their energy? What are the collision cross sections? It is in this direction that we are

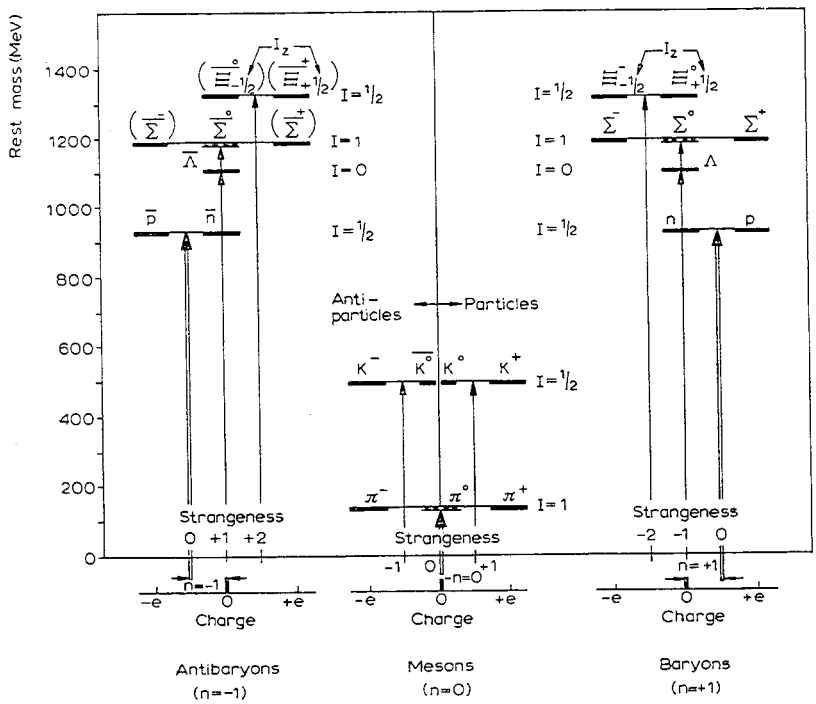


Fig. 2. A diagram showing all strongly interacting particles as known or predicted today. The particles still unobserved are in parenthesis. The weak interacting particles not reported in this diagram are the $\mu^{\pm}$ meson, the electron and positron, the neutrino and antineutrino, and the light quanta. (From Gell-Mann and Rosenfeld, *Ann. Rev. Nucl. Sci.*, 7 (1957) 407.)

working now and here we must be guided mainly by experiment, at least for the time being, and also be prepared for surprises.

The first surprise came immediately after the discovery of the antiproton when we found that this particle has an unusually large collision cross section. This fact has now been studied intensively for some time. The simplest situation occurs in the case of proton–antiproton collisions. There, in addition to the charge exchange process mentioned above, there are two other possibilities, elastic scattering and annihilation, at least until we reach energies such that inelastic processes (pion production) also become possible. Thus we have three cross sections: for scattering, for annihilation, and for charge exchange. All three have been measured for a wide energy interval and the results are shown in Fig. 3.

The magnitude of these cross sections is striking when we compare them

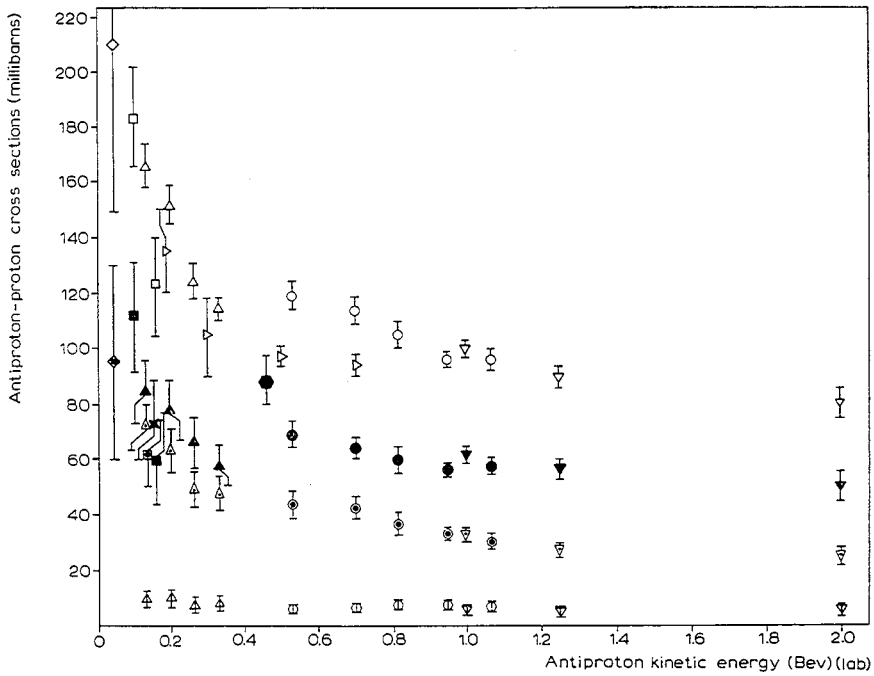


Fig. 3. All $\bar{p}$-p cross sections published up to November, 1959. The *open symbols* are total cross sections; *closed symbols* are inelastic cross sections (which are due to annihilation only for $T_{\bar{p}} \lesssim 290$ MeV); *open symbols encircling a dot* are elastic cross sections; *open symbols crossed by a vertical line* at the bottom of the figure are charge exchange cross sections.

The various symbols are referenced as follows:

□ Agnew, Elioff, Fowler, Gilly, Lander, Oswald, Powell, Segrè, Steiner, White, Wiegand, Ypsilantis, *Bull. Am. Phys. Soc.*, Ser. II, 4 (1959) 357.

▽ Armenteros, Coombes, Cork, Lambertson, Wenzel, *Bull. Am. Phys. Soc.*, Ser. II, 4 (1959) 356.

⬡ Chamberlain, Keller, Mermod, Segrè, Steiner, Ypsilantis, *Phys. Rev.*, 108 (1957) 1553.

△ Coombes, Cork, Galbraith, Lambertson, Wenzel, *Phys. Rev.*, 112 (1958) 1303.

○ Elioff, Agnew, Chamberlain, Steiner, Wiegand, Ypsilantis, *Phys. Rev. Letters*, 3 (1959) 285.

▷ Cork, Lambertson, Piccioni, Wenzel, *Phys. Rev.*, 107 (1957) 248.

◇ Horwitz, Miller, Murray, Tripp, *Phys. Rev.*, 115 (1959) 472.

★ Emulsion results of many authors compiled and averaged by Baroni *et al.*, *Nuovo Cimento*, 12 (1959) 564.

with those obtained in proton–proton collisions. A tentative theory of this phenomenon has been put forward by Chew[9] and his associates, and also by Koba and Takeda in Japan[10].

The model is based on the Yukawa theory of nuclear interactions in such a way as to stress the analogy between the nucleon–nucleon and the nucleon–antinucleon system. For the nucleon–nucleon system a model consisting of a hard repulsive core of a radius of about $\frac{1}{3}$ of the compton wavelength of the pion (0.45×10^{-13} cm) surrounded by a pion cloud has been reasonably successful in explaining the experimental results of the scattering and polarization experiments. The pion cloud which dominates the interactions at moderate distance can be treated from first principles of pion theory. The hard repulsive core on the other hand is unaccounted for from a pion theoretical point of view and must be introduced *ad hoc* as a phenomenological hypothesis, although the existence of heavier mesons such as the K mesons may have something to do with it. For a nucleon–antinucleon system the pion cloud of the antinucleon is substituted by its charge conjugate according to the expectations of meson theory and the medium range interactions are treated on the basis of this theory. The overlap of the cores, however, is now supposed to bring annihilation instead of strong repulsion. On the basis of this model it has been possible to account for most of the observations made thus far, which however do not extend to energies above 1 BeV where some critical tests of the theory await us.

In addition to the total cross sections for scattering, annihilation and charge exchange mentioned above, the angular distribution on scattering has been measured. Here a large diffraction peak in the forward direction has been found. It is directly related to the annihilation.

The extension of the cross section studies to complex nuclei has been started. The deuteron has been first investigated with the hope of finding information on the neutron–antiproton interaction. Here the data are still very rough, mainly because the subtraction techniques which we were forced to use introduce considerable errors. The qualitative feature seems to be that there is not much difference between proton–antiproton and neutron–antiproton collisions.

For heavier nuclei the data from the nucleon–antinucleon collision have been fed into an optical model treatment and the results agree with the experimental data as far as they are available. This gives a consistent picture connecting the more complicated case to the simpler one.

There are however still some crucial tests to be performed on the p–$\bar{p}$ case

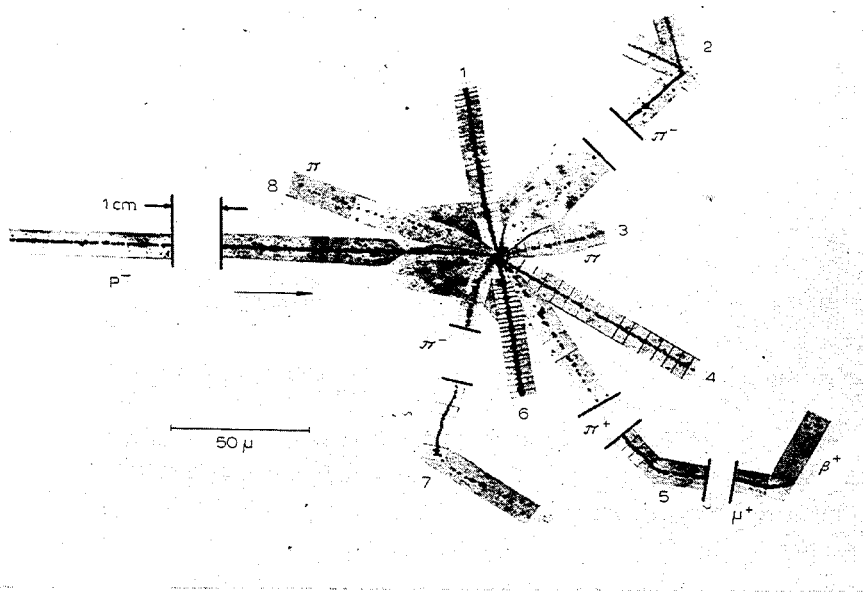


Fig. 4. An annihilation star showing the particles as numbered.

No.	1	2	3	4	5	6	7	8
Identity	p?	π^-	π?	p	π^+	H^3(?)	π^-	π
T (MeV)	10	43	175	70	30	82	34	125

Total visible energy 1,300 MeV. Total energy release > 1,400 MeV.

in order to validate the Chew model. At high energy, say 2 BeV, the annihilation cross section should be essentially the cross section of the core, and hence considerably smaller than the one observed at lower energy: 3×10^{-26} cm^2 would be a generous guess. If this expectation is not fulfilled it will be necessary to look for some other model. I will not go further into the numerous problems connected with cross-section studies, and will turn now to the annihilation.

The annihilation process itself has been fairly well investigated experimentally, but the theoretical situation leaves much to be desired. Initially the effort was mainly directed toward establishing the fact that the energy released was $2mc^2$ (m is the mass of the proton, c the velocity of light), thus furnishing a final proof of the annihilation. In the early investigations with photographic emulsions carried out in my group especially by Gerson Goldhaber and by a group in Rome led by Amaldi, we soon found stars showing

a visible energy larger than mc^2, giving conclusive evidence of the annihilation in pairs of proton and antiproton[11]. With great pleasure I recognized in the Nobel diploma the image of the first star of this type, found in Berkeley by Prof. Gösta Ekspong, now of Stockholm. It is shown in Fig. 4.

The observations on annihilation have been performed with many techniques. Initially, immediately after the identification of the antiproton, these particles were stopped in a block of heavy glass and the showers due to the gamma rays resulting from the decay of neutral pions were observed by Dr. Moyer and his co-workers[12]. This method was not however very quantitative.

Photographic emulsions were also exposed to antiprotons at the earliest possible moment. Here we see only the charged annihilation products, although much detailed information is obtainable. The great observational effort needed here was shared in a large cooperative experiment in which many laboratories in the U.S.A. and in Europe participated[13].

Bubble chambers have also been used, both of the propane and of the hydrogen type.

By now we know a good deal about annihilation. It gives rise prevalently to π mesons. These in a time of the order of 10^{-8} seconds decay into μ mesons and neutrinos. The μ mesons in a time of the order of microseconds decay into electrons or positrons and neutrinos, and the electrons and positrons finally recombine to give gamma rays. In a few microseconds the total rest mass of the nucleon–antinucleon pair degrades to particles with rest mass zero, travelling away from the spot of the annihilation with the velocity of light.

Direct annihilation into photons may occur, but is expected to be rare and thus far has never been observed with certainty.

The reason for this difference between the behavior of electron–positron and nucleon–antinucleon pairs is, of course, that the latter can annihilate not only through the electromagnetic interaction giving rise to light quanta, but also through the specific nuclear interaction whose quanta are the pions. This last interaction is much stronger than the electromagnetic one and when both are simultaneously present its effects overwhelm those of the electromagnetic interaction, which is the only available to the electron–positron pair.

The most significant result of the annihilation studies is that the annihilation process gives rise to an average of 4.8 pions per annihilation, about equally divided among positive, negative, and neutral pions. These pions

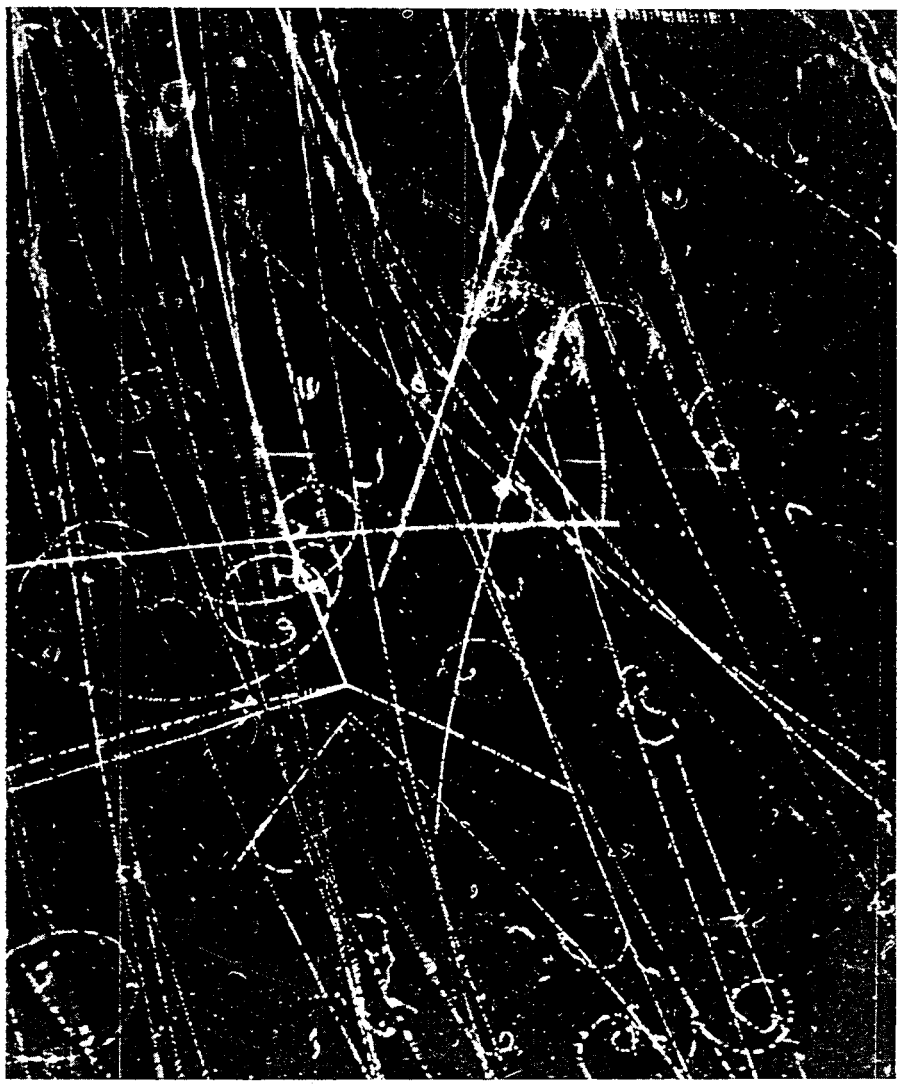


Fig. 5. Annihilation of an antiproton in carbon giving rise to a K meson and a Λ hyperon.

escape with a continuous energy distribution, the average kinetic energy being about 200 MeV. In about 4 percent of the annihilation cases at rest strange particles, K mesons, are emitted. See Fig. 5.

The escaping pions give rise in complex nuclei to secondary processes and thus a number of nucleons or light nuclei is also found among the particles

emitted on annihilation. Sometimes the relatively rare K mesons interact producing a Λ hyperon and even more complicated hyperfragments have been observed (Ekspong).

In hydrogen the multiplicity of the prongs, referring of course only to charged particles, is given in the following little table, for annihilations at rest. Naturally only even numbers of charged prongs may appear because the total charge of the proton–antiproton system is zero. (See Table 2.)

Table 2.

Charged multiplicity	0	2	4	6	8
Numbers of stars (total, 222)	10	89	109	14	0

From the theoretical point of view, we do not yet have an entirely satisfactory picture of the annihilation process. It has been mostly analyzed on the basis of a statistical theory put forward many years ago by Fermi, which does not take into account any detailed mechanism, but only the obvious and necessary features determined by phase space. This theory contains only one free parameter, namely, the volume into which the energy released on annihilation is concentrated at the beginning of the phenomenon. Naturally this volume is supposed to be the one corresponding to a sphere of the radius equal to radius of action of nuclear forces. If one calculates what is to be expected on this basis, one finds a result which is in rather poor agreement with experiment, namely, the multiplicity of pions produced is larger than that predicted by the model. Clearly the average energy and the multiplicity are connected, and hence the average energy also disagrees with the naive statistical prediction. The model can be made to yield correct results by increasing beyond what seems plausible the volume in which the energy comes to equilibrium. Many attempts have been made to refine Fermi's theory and to bring it into agreement with facts. Some of these attempts are very ingenious and one would wish that there were more success than there is. The ratio between K mesons and pions is another element of the puzzle that has to be taken into account and seems rather intractable for the time being.

It is however hardly to be expected that a purely statistical theory should explain quantitatively the annihilation process, inasmuch as selection rules, strong interactions of the escaping particles and other important factors completely omitted in the theoretical picture, are at work. I think that the future study of the annihilation process with its bearing on the core of the nucleon,

a region of which we know so little, will give some important results. Antinucleons are especially suited for this study because they will exhibit more clearly than other particles the effects of the core.

And now let me say some words on the popular subject of the « antiworld ». Already Dirac in his Nobel Lecture of 1933 said:

« If we accept the view of complete symmetry between positive and negative electric charge so far as concerns the fundamental laws of nature, we must regard it rather as an accident that the earth (and presumably the whole solar system), contains a preponderance of negative electrons and positive protons. It is quite possible that for some of the stars it is the other way about, these stars being built up mainly of positrons and negative protons. In fact, there may be half the stars of each kind. The two kinds of stars would both show exactly the same spectra, and there would be no way of distinguishing them by present astronomical methods. »

We can now add that the proved existence of the antinucleons has very strongly corroborated this possibility, although we also know that the symmetry between electric charges breaks down for weak interactions. As far as astronomical means are concerned, a verification seems impossible in principle, because they depend on electromagnetic phenomena, which are invariant under charge conjugation. It is however interesting that the recent important discoveries about β-decay and the neutrino now give a method which, while still impossible in practice, is sound in principle, being based on weak interactions which are *not* invariant under charge conjugation. This method, if it could be executed, would solve unambiguously the question of the existence of antiworlds. If we observe a star and from its astronomical characteristics can decide that most of its energy comes from a known cycle, as for example the carbon cycle, which is dominated by β-decays, we can see whether the antineutrinos coming from it are or are not of the same kind as the antineutrinos coming from a pile or from our sun by performing an inverse β-decay experiment. If it should turn out that they are neutrinos, i.e. different from those coming from the sun, then the star is of antimatter.

Let me finish this lecture with a remark and some acknowledgements. As in many investigations in high energy physics in recent times, this experiment is the result of a large cooperative effort. The credit for the success is shared by many individuals and even by a machine, which was obviously necessary to produce particles above the threshold for nucleon pair production. Since it is impossible to mention all the numerous contributors, I shall limit myself to a few. Dr. O. Piccioni helped materially in the early plan-

ning of the experiment, especially by suggesting the use of magnetic quadrupole lenses. Dr. E. Lofgren most ably directed the operation of the Bevatron. Dr. H. Steiner supplied invaluable help during the whole experiment. Dr. T. J. Ypsilantis, our colleague and co-author, also worked with us all the time. Above all, however, our co-author and comrade of 20 years of work, Dr. Clyde Wiegand, was indispensable and deserves a major part of the credit for the success of our investigation.

1. See for instance: P. A. M. Dirac, *Les Prix Nobel en 1933.*
2. O. Chamberlain, E. Segrè, C. Wiegand, and T. Ypsilantis, *Phys. Rev.*, 100 (1955) 947.
3. O. Chamberlain, W. W. Chupp, G. Goldhaber, E. Segrè, C. Wiegand; and E. Amaldi, G. Baroni, C. Castagnoli, C. Franzinetti, A. Manfredini, *Phys. Rev.*, 101 (1956) 909.
4. O. Chamberlain, E. Segrè, C. Wiegand, and T. Ypsilantis, *Nature*, 177 (1956) 11.
5. B. Cork, G. R. Lambertson, O. Piccioni, and W. A. Wenzel, *Phys. Rev.*, 104 (1956) 1193.
6. L. E. Agnew, T. Elioff, W. B. Fowler, L. Gilly, R. Lander, L. Oswald, W. Powell, E. Segrè, H. Steiner, H. White, C. Wiegand, and T. Ypsilantis, *Phys. Rev.*, 110 (1958) 994.
7. M. Baldo-Ceolin and D. J. Prowse, *Bull. Am. Phys. Soc.*, 3 (1958) 163.
8. Button, Eberhard, Kalbfleisch, Lannutti, Maglić, Stevenson, *Phys. Rev. Letters* (to be published).
9. J. S. Ball, and G. F. Chew, *Phys. Rev.*, 109 (1958) 1385.
10. Z. Koba, and G. Takeda, *Progr. Theoret. Phys. (Kyoto)*, 19 (1958) 269.
11. O. Chamberlain, W. W. Chupp, A. G. Ekspong, G. Goldhaber, S. Goldhaber, E. J. Lofgren, E. Segrè, C. Wiegand; and E. Amaldi, G. Baroni, C. Castagnoli, C. Franzinetti, A. Manfredini, *Phys. Rev.*, 102 (1956) 921.
12. J. M. Brabant, B. Cork, N. Horwitz, B. J. Moyer, J. J. Murray, R. Wallace, and W. A. Wenzel, *Phys. Rev.*, 101 (1956) 498.
13. W. H. Barkas, R. W. Birge, W. W. Chupp, A. G. Ekspong, G. Goldhaber, S. Goldhaber, H. H. Heckman, D. H. Perkins, J. Sandweiss, E. Segrè, F. M. Smith, D. H. Stork, L. van Rossum; and E. Amaldi, G. Baroni, C. Castagnoli, C. Franzinetti, A. Manfredini, *Phys. Rev.*, 105 (1957) 1037.

Biography

Emilio Segrè was born in Tivoli, Rome, on February 1st, 1905, as the son of Giuseppe Segrè, industrialist, and Amelia Treves. He went to school in Tivoli and Rome, and entered the University of Rome as a student of engineering in 1922. In 1927 he changed over to physics and took his doctor's degree in 1928 under Professor Enrico Fermi, the first one under the latter's sponsorship.

He served in the Italian Army in 1928 and 1929, and entered the University of Rome as assistant to Professor Corbino in 1929. In 1930 he had a Rockefeller Foundation Fellowship and worked with Professor Otto Stern at Hamburg, Germany, and Professor Pieter Zeeman at Amsterdam, Holland. In 1932 he returned to Italy and was appointed Assistant Professor at the University of Rome, working continuously with Professor Fermi and others. In 1936 he was appointed Director of the Physics Laboratory at the University of Palermo, where he remained until 1938.

In 1938 Professor Segrè came to Berkeley, California, first as a research associate in the Radiation Laboratory and later as a lecturer in the Physics Department. From 1943 to 1946 he was a group leader in the Los Alamos Laboratory of the Manhattan Project. In 1946 he returned to the University of California at Berkeley as a Professor of Physics, and still occupies that position.

The work of Professor Segrè has been mainly in atomic and nuclear physics. In the first field he worked in atomic spectroscopy, making contributions to the spectroscopy of forbidden lines and the study of the Zeeman effect. Except for a short interlude on molecular beams, all his work until 1934 was in atomic spectroscopy. In 1934 he started the work in nuclear physics by collaborating with Professor Fermi on neutron research. He participated in the discovery of slow neutrons and in the pioneer neutron work carried on in Rome 1934–1935. Later he was interested in radiochemistry and discovered together with Professor Perrier the element technetium, together with Corson and Mackenzie the element astatine, and together with Kennedy, Seaborg, and Wahl, plutonium-239 and its fission properties.

His other investigations in nuclear physics cover many subjects, e.g., isomerism, spontaneous fission, and lately high-energy physics. Here he, his associates and students have made contributions to the study of the interaction between nucleons and on the related polarization phenomena. In 1955 together with Chamberlain, Wiegand, and Ypsilantis he discovered the antiproton. The study of antinucleons is now his major subject of research.

Professor Segrè has taught in temporary appointments at Columbia University, New York, at the University of Illinois, at the University of Rio de Janeiro and in several other institutions. He is a member of the National Academy of Sciences (U.S.A), of the Academy of Sciences at Heidelberg (Germany), of the Accademia Nazionale dei Lincei of Italy, and of other learned societies. He has received the Hofmann Medal of the German Chemical Society and the Cannizzaro Medal of the Italian Accademia dei Lincei. He is an Honorary Professor of San Marcos University in Peru and has an honorary doctor's degree of the University of Palermo, Italy.

Professor Segrè is married to Elfriede Spiro; they have a son, Claudio, and two daughters, Amelia and Fausta.

Physics 1960

DONALD ARTHUR GLASER

«for the invention of the bubble chamber»

Physics 1960

Presentation Speech by Professor K. Siegbahn, member of the Swedish Academy of Sciences

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

Surely we all have often admired the beautiful white streaks which are left against the blue heaven by a highflying jet airplane. These streaks are made up of very small, finely divided waterdrops which have been condensed into a cloud track behind the plane. Long after the plane has vanished in the distance, one may in detail, with the help of these cloud tracks, trace every movement made by the plane. A similar visual method is utilized in nuclear physics in order to expose the passing of individual atomic particles through a gas. It is the famous « cloud chamber » which has played such an important role in nuclear physics and which provided its English inventor, C. T. R. Wilson, the Nobel Prize in Physics 33 years ago, that is, 1927. There prevails a certain connection between the invention which is today awarded this prize and Wilson's invention; we therefore have reason to remind ourselves of the Wilson chamber's method of function: An atomic particle, such as one from a radioactive material, produces along its path charged fragments which are called ions. If the atomic particle is made to pass through a gas chamber which contains a supersaturation of water vapor, the ions will function as centers of condensation, and small water-droplets are formed along the particle's path, just as in the case of the jet-planes' vapor streak. The supersaturation of water vapor was brought about at a suitable chosen instant by allowing the gas in the chamber to rapidly expand with the help of a mobile piston in the chamber's base. If simultaneously there is a light flash illuminating the gas in the chamber, the fine cloud-track formed by the chamber's expansion after the particle's passage can be photographed. One may photograph such processes as atomic decay, nuclear fission, and cosmic rays, just to mention a few examples. It is truly a wonderful invention, especially if one considers the fact that the atomic particles which are made visible in this refined manner are no larger than one-hundredth of a billionth of a millimeter.

The Wilson chamber has certainly played a tremendously important role, especially during the 30's, which is referred to as « the golden age of nuclear

physics», and there is no doubt that it was the Wilson chamber which made possible the greatest nuclear physical discoveries during that decade. It was also an ideal research instrument for that time, because the nuclear particles which were then of interest, and which were possible to depict by artificial means twenty to thirty years ago, had ranges which, for the most part, are of the order of several decimeters in a gas at normal pressure. In other words, particles had energies of some million volts and there was no difficulty in using the Wilson chambers, which were sufficiently large, in order to be able to photograph paths of all incoming particles along its entire length as well as all of the possible nuclear reactions which would have enough room to occur within the Wilson chamber.

The situation is completely different in the nuclear physics of today, where one now has at one's disposal particle accelerators with energies as high as 25 billion volts, for example the accelerator which has recently been built at the European nuclear research center in Geneva. In other words, energies which are more than 1,000 times larger than those which were earlier obtainable. In order to be able to register such particles during the entire time of flight interval, one would be forced to use a Wilson chamber of the inconceivable size of 100 meters or more. It is obviously necessary to use a medium other than gas in order to be able to bring such particles to rest. Donald Glaser has succeeded in solving this problem, and his so-called «bubble chamber» is the *high-energy nuclear physics* counterpart to the *low-energy nuclear physics* Wilson chamber. As you will remember, it was last year that the discovery of the so-called «antiparticles» was rewarded with the Nobel Prize. An anti-particle is the «inversion» of a particle, its «mirror image». One could say that Glaser's bubble chamber is an anti-Wilson chamber. Particle tracks in Glaser's chamber are composed of small *gas bubbles in a liquid.*

We depart from this and the jet plane now, and turn our attention instead to something as prosaic as what happens when one opens a soda-water bottle. In this manner we follow with all sureness Donald Glaser's own train of thought from the moment he received his first impulse toward his invention of the bubble chamber. If we ease up on the pressure in the bottle by removing the cap, bubbles will rise from the liquid. The bubbles come preferably from certain points which serve as centers for bubble development. Glaser's next step in his line of reasoning was to use, instead of a soda-water bottle, a liquid which is heated to a temperature near its boiling point and which is enclosed in a chamber. If the pressure is rapidly eased from above

the liquid, for example by use of a mobile piston, the liquid will have a tendency to boil. With a certain amount of care, one is able to maintain the liquid in this superheated, unstable state without boiling. The slightest disturbance in the liquid, however, gives rise to an instantaneous boiling, exactly as did the opening of a soda-water bottle. Glaser's idea was that an atomic particle passing through the liquid would be able to provoke boiling by means of the ions which the atomic particle produces along its path and which act as bubble-development centers. If a flash picture were subsequently taken of the superheated liquid immediately after an atomic particle passes through, one ought to be able to observe the particle's path which is followed by the small bubbles which would have been produced. It is necessary, of course, that one proceeds quickly, for immediately afterwards, the bubble-track will have degenerated into a general boiling of the liquid. This was Glaser's sound idea and, working according to his plans in a systematic manner, he was successful in realizing the first radiation-sensitive bubble chamber in 1952. Even if the principle of Glaser's bubble chamber can be considered simple, it represents an exceedingly difficult development program requiring several years of work which lie behind the completed invention which will now be awarded this prize.

Rather soon after Glaser had published his ideas and the results of his first experiments, there were several persons who realized that something important would come of this. Several other scientists also left important contributions to the practical shaping of different types of bubble chambers, but Glaser is the one who made the really fundamental contributions. In order to get his apparatus to function, Glaser was forced to consider the physics of bubble formation both from theoretical and experimental points of view. As usual, it turned out that only a systematic procedure for studying the complete problem led to a solution.

The most striking feature of development during the most recent years is without doubt the increase in size of the bubble chamber. Glaser's first little glass-receptacle of some centimeters in size and filled with ordinary ether has successively grown to extraordinary voluminous proportions which represent the engineering-art's most exclusive subtleties. The largest chamber built to the present time is close to 2 m long, 0.5 m wide and deep, and contains liquid hydrogen which is condensed by a large cooling device providing temperatures in the vicinity of absolute zero. This largest liquid chamber is surrounded by a powerful electromagnet which is capable of bending the paths of the particles so that the faint bubble-tracks become

slightly curved. In this way one is able to identify the unknown atomic particles when they, travelling very close to the speed of light, pass through the chamber. The large bubble chamber has also an extremely complicated automatic read-off and calculation apparatus which sends information from the tracks in the bubble chamber into a larger mathematical-computer, which, in turn, after a moment's thought, forwards from the world of atoms the news which the nuclear researcher so eagerly awaits. This part of the set-up has received the characterization name of « Frankenstein ».

By using Glaser's bubble chamber the modern nuclear researcher has at his disposal just the scientific instrument which is required in order to exploit the gigantic atomic accelerators which in recent years have been constructed in atomic research centers in the U.S.A., West Europe, and Russia. Large research teams are now at work investigating the strange, new particles which are formed, transformed, and annihilated when the beam from these machines is directed into the bubble chamber; how atomic nuclei are split, and how from the atomic fragments new particles are again generated to later change guise and finally be destroyed.

Dr. Glaser. Your invention of the « bubble chamber » has opened up a new world for nuclear science. We can now observe with our own eyes all the strange processes which occur when high-energy beams from BeV machines are directed into your chamber. Already a great amount of information has been obtained in this way and many important discoveries will no doubt follow in the near future by means of your method. It is unusual for a development in modern nuclear physics to be due to such a large extent to one single man.

On behalf of the Royal Academy of Sciences I congratulate you most heartily. May I now ask you to receive the Nobel Prize from the hands of His Majesty the King.

DONALD A. GLASER

Elementary particles and bubble chambers

Nobel Lecture, December 12, 1960

It is a very great honor and privilege that I have this opportunity to describe to you the invention, development, and uses of bubble chambers for studying elementary particles.

From the earliest beginning of serious scientific thought up to the present day, men have tried to understand the properties of matter by imagining it to be built up out of a small number of basic irreducible elements. According to current scientific ideas, these irreducible elements are actually submicroscopic particles which are supposed to be indivisible and therefore not made up out of yet smaller particles. We imagine all matter to be agglomerations of molecules built up from atoms which are, in turn, constructed from electrons, protons, and neutrons. These three particles have definite masses, spins, electric charges, magnetic moments, and other properties. All electrons are supposed to be identical and so are all protons and all neutrons. During the last twenty years a number of other apparently indivisible particles have been discovered so that now there are thirty of them altogether. Although a few of these thirty have not yet been observed experimentally, they are included in the list because the theoretical expectation of their existence is very strong and it seems virtually certain that they will be observed during the next few years. Some of the thirty elementary particles were found experimentally, only after theoretical arguments had predicted their existence and the conditions required for creating and observing them. Others were discovered more or less accidentally before we had any theoretical ideas about them or their properties. Among these are the μ-meson and the first « strange particles ». Although we have a lot of experimental knowledge about them there is still no satisfactory mathematical theory which can predict their physical properties or interpret their existence in terms of the other known elementary particles. Physicists find it distasteful to believe that all thirty of these elementary particles are « really fundamental », and hope some day to understand them all as compounds of a smaller number of truly fundamental particles or in terms of some other concepts involving simple underlying laws.

Advances in our knowledge of the elementary particles have depended heavily on the development of techniques for detecting them and observing their properties in detail. Among the most important instruments used in these experimental studies have been the cloud chambers and the nuclear emulsion. Both techniques permit detailed visual study of the paths of charged elementary particles, but each has limitations that seriously hamper some studies of elementary particles and other phenomena of high-energy nuclear physics.

I became interested in trying to devise new experimental methods for investigating the physics of elementary particles in 1950, not long after the new « strange particles » had been discovered in cosmic rays. In those days a rather small number of these particles had been observed, and they were still called « V-particles » or « pothooks » because of their unusual appearance in the cloud chamber photographs. In fact, I remember that when I left the California Institute of Technology in 1949 after finishing my doctoral research on cosmic radiation under the direction of Professor Carl D. Anderson, there was written at the top of his blackboard the question: « What have we done about the pothooks today? » No one had predicted the existence of these particles or had any idea how they might fit into theoretical schemes describing the particles already well known at that time. Their discovery in cosmic ray interactions in 1947 by Butler and Rochester and later observation by others created high excitement among physicists. Here was a whole new family of particles that might well lead to some novel and deep ideas about the laws and symmetries of Nature on the submicroscopic level of the elementary particles. More experimental information was desperately needed concerning the production, decay, and interactions of these new particles.

Greatly stimulated by these developments, I began to wonder whether it would be possible somehow to speed up the rate of observing the strange particles and their interactions. Most examples of their decays were seen in cloud chamber photographs of penetrating cosmic ray showers originating in lead plates just above the chamber. Because of the rarity of strange-particle production and the low density of gas in a cloud chamber, production events almost never occurred in the gas itself. For this reason it was difficult to get convincing experimental evidence on even such simple ideas as the law of associated production, which had been invented to explain how the new particles could have such long lifetimes. Nuclear emulsions are not useful in these studies because the neutral « V particles » have typical mean decay

lengths of a few centimeters. This makes it virtually impossible to associate correctly a decaying Λ° or K° with its parent star in an emulsion stack containing a reasonable density of events.

There was therefore a great need for a particle detector of high density and large volume – tens to hundreds of liters – in which tracks could be photographed and scanned at a glance, and in which precision measurements of track geometry could be made. Expansion cloud chambers which operated with internal gas pressures up to 300 atmospheres had been built for this reason, but they were elaborate and cumbersome machines which required waiting times of 15 to 30 minutes between expansion cycles.

Since all detectors capable of recording the passage of a single minimum ionizing particle must use some amplifying process to be sensitive to the minute amounts of energy deposited by a fast charged particle, I began to list all the amplifying mechanisms I could imagine that might serve as the basis for a detector of the type that was needed. Previous detectors had used the instability of a gas, liquid, or solid, against dielectric breakdown in an intense electric field, the chemical instability of nuclear emulsion with respect to its developing solution, or the thermodynamic instability of a supercooled or supersaturated vapor. Since I wished to attain high density without using very high pressures, I considered mainly instabilities that can exist in liquids and solids, such as chemical instabilities including the tendency of monomers to polymerize, instabilities due to intense electric fields, and thermodynamic instabilities such as are present in supercooled liquids, superheated solids, or superheated liquids. At the time that I was studying these instabilities, I knew that the large proton synchrotons in the few GeV energy range would come into operation in the early 1950's and that they would have pulse repetition times of a few seconds. It was therefore important that the new detector be able to cycle in a few seconds to be most useful with these machines as well as with the cosmic rays. For this reason I rejected chemical and solid systems as being probably too difficult to recycle rapidly.

The thermodynamic instability of a superheated liquid can be used to detect minimum ionizing radiation only if the density of ionization energy deposited in the liquid along the path of the particle is sufficient to form a vapor bubble nucleus large enough to grow to photographable size. If the vapor pressure in a bubble at the temperature T exceeds the external pressure applied to the liquid by the amount ΔP, the radius of this critical nucleus is

$$r_c = \frac{2\sigma(T)}{\Delta P}$$

where $\sigma(T)$ is the surface tension of the liquid–vapor interface. A lower limit for the nucleation energy required for forming this bubble is obtained by calculating the energy required for its isothermal reversible formation

$$W = \frac{16\pi\sigma^3(T)}{3\,(\Delta P)^2}$$

Then the minimum energy density required per unit length of track is

$$\frac{W}{2r_c} = \frac{4\pi}{3}\frac{\sigma^2(T)}{\Delta P}$$

If we imagine that we can quickly reduce the external pressure on the hot liquid almost to zero compared with the vapor pressure as a means of preparing the final superheated state, we approximate ΔP by $P(T)$, the saturated vapor pressure. Then the energy loss of the minimum ionizing particle must exceed

$$\frac{W}{2r_c} \simeq \frac{4\pi}{3}\frac{\sigma^2(T)}{P(T)}$$

When we insert the actual value of the energy loss of a minimum ionizing particle, we find that the only way to attain the low surface tension and high vapor pressure required by this equation is to raise the temperature of the liquid to the vicinity of its critical temperature, since the surface tension vanishes at the critical temperature. In fact, it turns out that the temperature must be about one-half to two-thirds of the way from the boiling point to the critical point for successful bubble chamber operation.

Closer estimates of the required degree of superheat can be made only by assuming specific mechanisms for conversion of ionization energy to bubble nucleation energy. We must first ask whether macroscopically measured values of $\sigma(T)$ and $P(T)$ apply for bubbles 10^{-6} cm in diameter. In the theory of cloud chamber operation, it is important that the equilibrium vapor pressure at a convex liquid surface is much larger than at a flat surface. Is there a similar effect at concave surfaces? It was several weeks before I

could answer this question, because each thermodynamic argument I used led to the conclusion that the vapor pressure inside a vapor bubble is the same as that at a flat interface, though I found that nearly all of the thermodynamics textbooks disagreed with me. Finally I found a discussion of this question, which had apparently been a subject of a serious dispute as late as 1939, in an excellent book, *Kinetik der Phasenbildung*, by M. Volmer, in which my conclusions were supported.

Little experimental information on the surface tension of highly concave surfaces was available, but it seemed plausible that there should not be a large dependence on curvature when the radius of curvature is very much larger than the average intermolecular distance in the liquid.

Two detailed mechanisms for bubble formation suggested themselves. In one, it was supposed that clusters of ions of like sign are produced occasionally along the track, and form bubbles by mutual electrostatic repulsion. In the second, it was supposed that excited atoms and molecules formed directly by the primary particle and by ion recombination converted their excitation energy into local heating of the liquid through superelastic collisions of the second kind. Both models made bubble formation along the tracks of minimum ionizing particles appear to be a very difficult and implausible process. I was encouraged by the idea that even if the primary ionization energy was insufficient, secondary delta rays with even a few hundred volts of energy or more would form frequently along the track and deposit their energy densely in a small volume as they stopped in the liquid. Coulomb scattering would tend to curl up the stopping electron tracks and increase the volume density of their energy deposit.

For a first experimental test of these ideas, I chose diethyl ether because of its relatively low surface tension, critical temperature, and critical pressure, and because it was cheap and easy to obtain in pure form. One particular calculation using the electrostatic model predicted that diethyl ether would be nucleated to boil by ionizing radiation at about 140°C at one atmosphere although its boiling point is only 34.6°C. It seemed to me unreasonable that such an enormous degree of superheat should be attainable experimentally, so I looked in the literature of the physical chemists to see what was known about the maximum attainable superheats. Finally I found a remarkable paper* describing attempts to superheat diethyl ether, a liquid which these authors had apparently chosen for the same reasons of convenience that dictated my choice. Imagine my excitement when I read that they had been

* F. B. Kenrick, C. S. Gilbert, and K. L. Wismer, *J. Phys. Chem.*, 28 (1924) 1927.

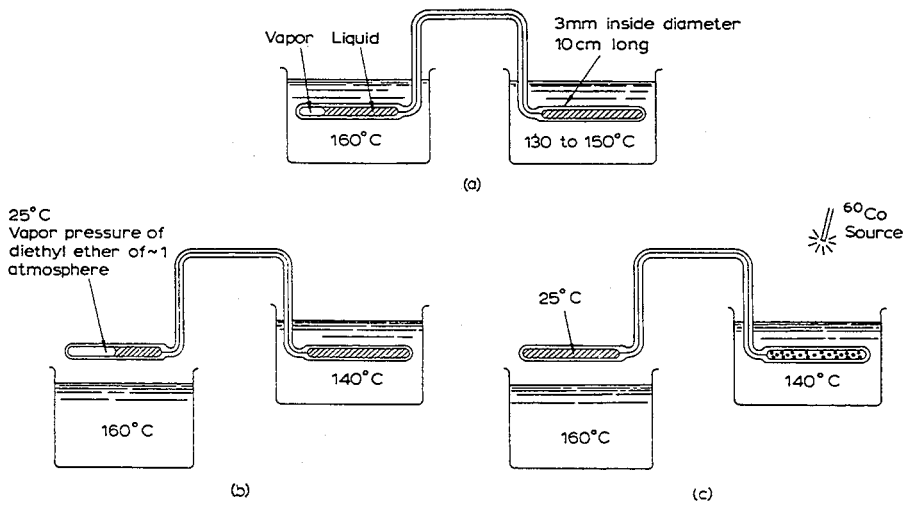


Fig. 1. Diagram of experiment demonstrating that superheated diethyl ether can be made to boil by ionizing radiation.
(*a*) Pressure on enclosed diethyl ether is high due to vapor pressure of the liquid at 160°C. Temperatures shown are maintained by hot oil baths; (*b*) 160°C bath removed. Temperature drops quickly in a few seconds to 25°C. Pressure on total system about 1 atmosphere. Diethyl ether in the bulb at 140°C is now highly superheated; (*c*) Boiling is triggered in the superheated diethyl ether by radiation from a ^{60}Co source.

able to maintain diethyl ether superheated for hours at 130°C and one atmosphere, but that at 140°C the liquid erupted at erratic time intervals after being brought rapidly to the high temperature. To demonstrate the « capriciousness » of this phenomenon, they quote a typical series of 30 consecutive « waiting times ». When I examined these times, I found them to be consistent with a Poisson distribution corresponding to the random occurrence of a nucleating event which distrupted their small volume of liquid with an average waiting time of about 60 seconds. From the reported geometry of the superheating apparatus, I estimated that its total « counting rate » for ionizing events in the liquid due to cosmic rays plus radioactive background at sea level was also about one count every 60 seconds!

To make a simple test of the hoped-for triggering of superheated diethyl ether by ionizing events, I did the experiment diagrammed in Fig. 1. A heavy-walled capillary tube of the shape shown was evacuated and filled with pure diethyl ether vapor and liquid. When the pressure on the liquid is reduced by cooling the bulb containing some vapor, the pressure drops

to about 1 atmosphere, superheating the liquid in the 140°C bulb. When exposed to gamma radiation from a ^{60}Co source, this hot ether erupts into violent boiling, instantly as far as the eye can tell.

It should be remarked parenthetically that before making these detailed calculations and experiments, I wanted to be sure not to overlook simple experimental possibilities, so I took some bottles of beer, ginger ale, and soda water into my laboratory, warmed them as much as I dared, and opened them with and without a radioactive source nearby. There was no apparent gross difference in the way they foamed. Water, of course, turns out to be just the wrong substance to use in a bubble chamber because it has a large surface tension and a high critical pressure.

Now that there was experimental proof of the reality of this new physical phenomenon on which a bubble chamber technology could be based, it was important to find out if minimum ionizing particles could initiate boiling and if the bubbles formed accurately along the path of the particle. Unless

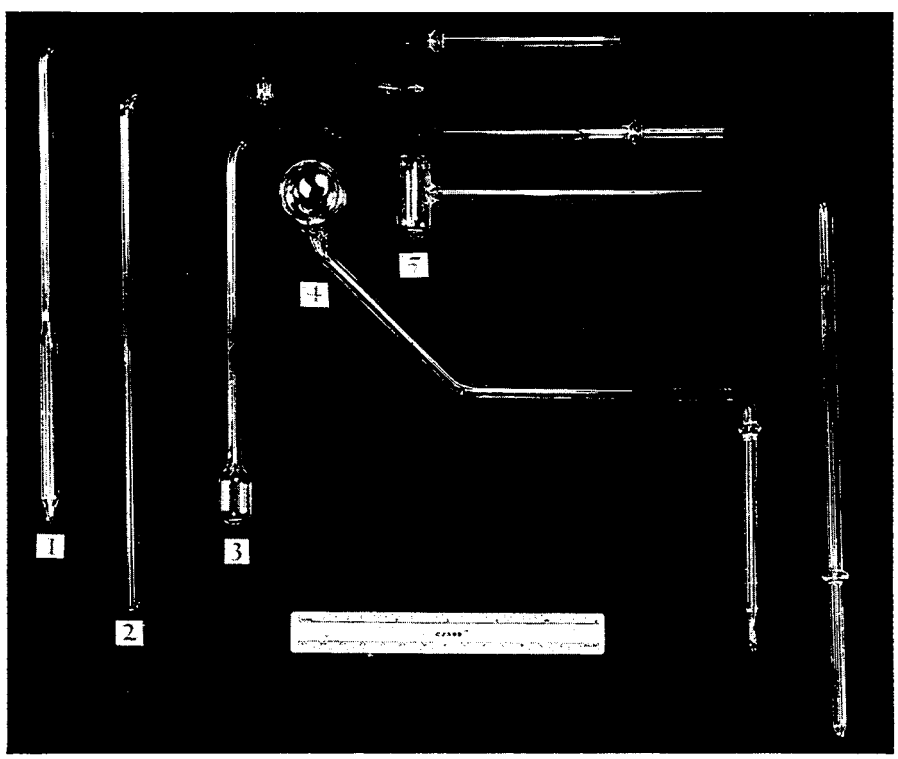


Fig. 2. Some of the first pyrex-glass bubble chambers.

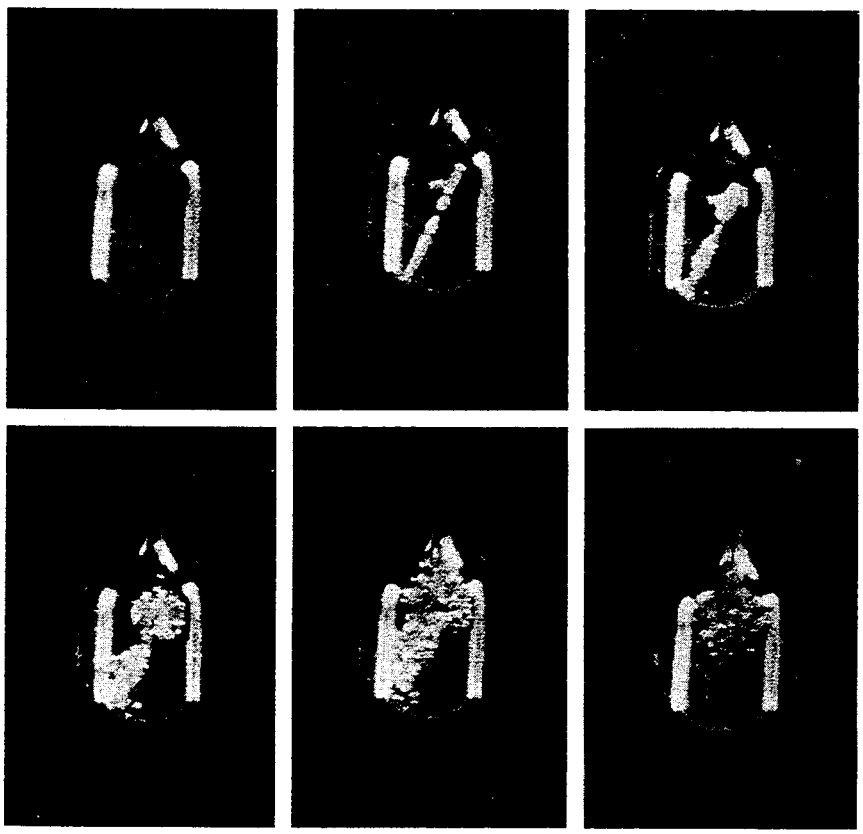


Fig. 3. High-speed movies taken at 3,000 pictures per second of eruptive boiling initiated in diethyl ether at 135°C by a penetrating particle. The pictures shown were taken at 0, $\frac{1}{3}$, 4, 22, 71, and 250 milliseconds. The first two are consecutive pictures. The chamber is about 1 cm inside diameter.

the bubble chamber could make accurate tracks of minimum ionizing particles, it would be of little use in high-energy physics experiments.

To look for tracks, I made a number of small pyrex-glass bubble chambers containing a few cm^3 each of diethyl ether. Some of them are shown in Fig. 2. A bath of hot oil maintained the required temperature, and a piston operated by a hand crank controlled the pressure. Each time the pressure was reduced by a quick turn of the hand crank, a high-speed movie camera recorded the onset of boiling after the usual few seconds of waiting time when the ether was in a quiescent superheated state. Sometimes the boiling

started along a well-defined track as shown in Fig. 3, which is taken using chamber 3 in Fig. 2. From these movies taken at 3,000 pictures/second, one sees that the bubbles grow to be more than a millimeter in diameter in 300 microseconds. Finer tracks were obtained by constructing an automatic device for expanding and recompressing the chamber every 10 seconds. Photographs of the chamber were taken with a xenon flashlamp whenever a vertical Geiger counter telescope indicated the passage of a penetrating cosmic ray particle during the few seconds of sensitive time following each expansion. These photographs, shown in Figs. 4 (*a*, *b*, *c*) proved that bubble chambers could yield precision measurements of events involving minimum ionizing tracks.

Although this series of experiments established with certainty that bubble chambers were feasible and had the right characteristics for elementary-particle studies, engineering development was needed before large chambers for serious experiments could be built. There were two main questions. What

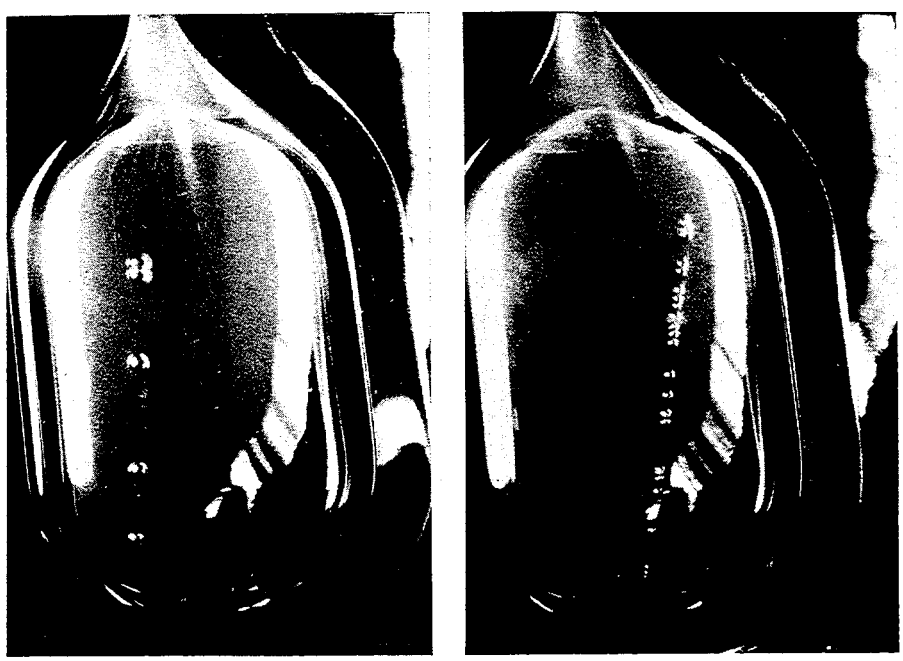


Fig. 4 (*a*, *b*). Penetrating cosmic ray tracks in 3 cm³ diethyl-ether bubble chamber. (Random expansion and counter-controlled flashlamps.)
(*a*) 60 microsecond flash duration, 139°C; (*b*) 10 microsecond flash delay, 20 microsecond flash duration, 140°C. (*Fig. 4 (c) on next page.*)

Fig. 4 (*c*). Penetrating cosmic ray tracks in 3 cm³ diethyl-ether bubble chamber. (Random expansion and counter-controlled flashlamps.)
(*c*) 10 microsecond flash delay, 5 microsecond flash duration, 141°C. The track in (*c*) is deflected about 2° in the middle of the chamber.

other liquids could be used in bubble chambers? How could chambers with hundreds of liters of sensitive volume be constructed? Many laboratories the world over pursued this development with great vigor and ingenuity.

Diethyl ether had been used in the first experiments for reasons of experimental convenience. Liquid hydrogen was obviously the physicists' dream as a working liquid because the interpretation of events in hydrogen are so straightforward. Using the electrostatic theory of bubble chamber operation, I estimated that hydrogen would work at about 27°K. Although I now have excellent reasons for believing the thermal or «hot track» theory of bubble nucleation is correct rather than the original electrostatic mechanism, the two theories give nearly the same scaling law for relating the operation conditions of different liquids. Therefore, the basically wrong electric theory gave correctly the operating temperatures for hydrogen, deuterium, helium, freon, propane, xenon, and a host of other liquids. At the University of Michigan there were no cryogenic facilities in 1953, so I trav-

elled to the University of Chicago and worked on liquid-hydrogen bubble chambers with Hildebrand and Nagle, who soon showed that superheated liquid hydrogen was radiation sensitive. Shortly after that, Wood at Berkeley photographed the first tracks in liquid hydrogen. Many other liquids were tested in our laboratory and in other places. No liquid that has been tested seriously has failed to work as a bubble chamber liquid. The choice of liquids depends only on the physical objectives of the experiment and on engineering and economic considerations.

It did not seem practical to make large chambers out of a single piece of pyrex glass. We therefore fabricated a small chamber with an aluminum body and flat glass windows sealed with rubber gaskets. Certainly there was no hope of maintaining a liquid superheated for any length of time in such a chamber because boiling begins easily at gaskets and scratches in the metal. We hoped that we could expand the chamber fast enough to keep the pressure in the interior of the chamber low for at least a few milliseconds, even though boiling begins instantly at the walls. Fig. 5 shows the 5-cm chamber used to test this method. Figs. 6 (*a*, *b*, *c*, *d*, *e*) shows a sequence of snapshots taken at various times after the beginning of the expansion. These pictures proved that a chamber fabricated of ordinary materials could have a sensitive time of a few milliseconds, long enough for use with large particle accelerators. Wood proved the same thing independently with his first small hydrogen chamber a short time before we did. These experiments made possible the design and construction of really large chambers containing hundreds of liters of liquid.

In Fig. 7 is shown a 15-cm propane chamber we exposed to the Brookhaven Cosmotron – the first bubble chamber to be used for experiments in high-energy physics. In Figs. 8, 9, 10, and 11 are shown our 30-cm propane chamber, 30-cm /21-liter/ xenon chamber, the xenon chamber fully assembled for an experiment, and finally the largest chamber operating now in the world, Alvarez's 180-cm /500-liter/ hydrogen chamber. Other really large chambers are in use or under construction at Brookhaven, Chicago, Geneva, Dubna, Saclay, London, and other high-energy physics centers. Smaller chambers in large numbers and great varieties are in use all over the world.

Large quantities of data on elementary particles and their interactions are being produced by these chambers. A number of new particles and phenomena have been discovered by their use. Precise information on masses, spins, lifetimes, parity violating decays, branching ratios, and polarizations has

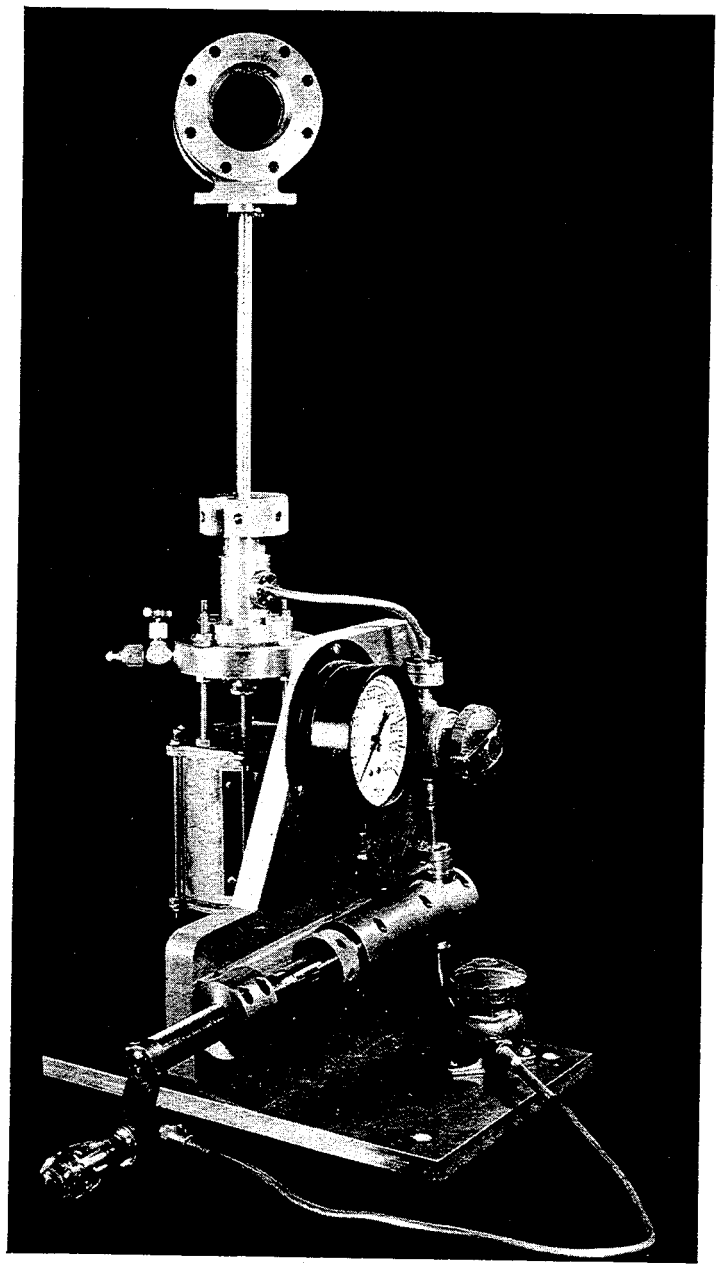


Fig. 5. An aluminum bubble chamber 5 cm in diameter with its expansion mechanism and volume adjustment system for testing the principle of the «dirty» bubble chamber.

been obtained for these particles. Nuclear fusion catalyzed by μ-mesons was discovered in a hydrogen bubble chamber. Figs. 12 through 14 are pictures showing some of the elementary-particle phenomena being studied now with these chambers.

Because of the high density and rapid cycling rate of bubble chambers, we now have abundant information on particle production, interaction, and decay as observed with beams at the large accelerators. It was a disappointment to discover that bubble chambers are not easy to use for cosmic ray experiments since they cannot be operated using counter-controlled expansions as can cloud chambers. We have established experimentally that the lifetime of the «latent bubble-track image» produced by a charged particle is less than 10^{-4} seconds and our theoretical estimates indicate that it may actually be about 10^{-8} seconds. Since mechanical expansion of a bubble

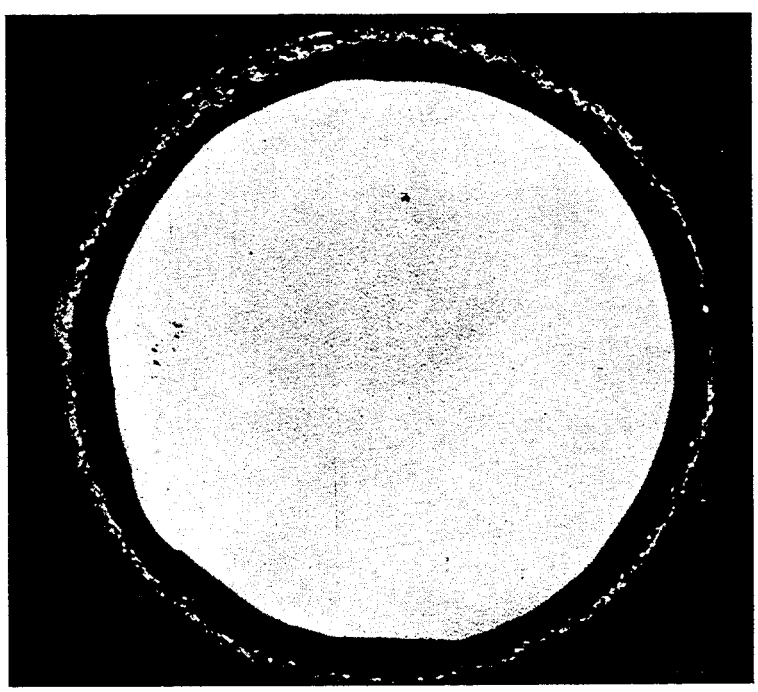


Fig. 6 (*a*). Operation of a «dirty» diethyl-ether chamber, 5 cm in diameter and 2.5 cm deep at 154°C. Photographs are taken with a 5 microsecond flash duration at different moments after the beginning of the expansion process; *11 milliseconds, no radiation:* violent boiling occurs at the gaskets but no bubbles have formed in the interior of the liquid or at the glass windows. (*Figs. 6 (b, c, d, e) on following pages.*)

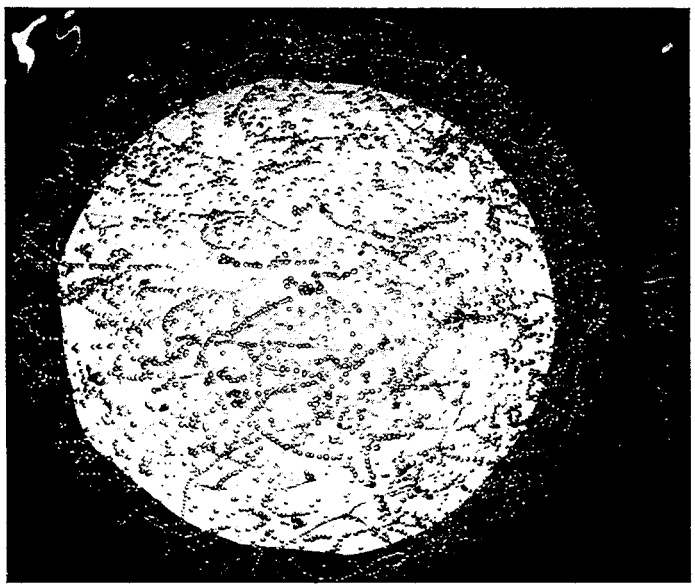


Fig. 6 (*d*). *7.55 milliseconds, radium source nearby:* although the bubbles on the oldest tracks have grown quite large, new tracks are still being formed.

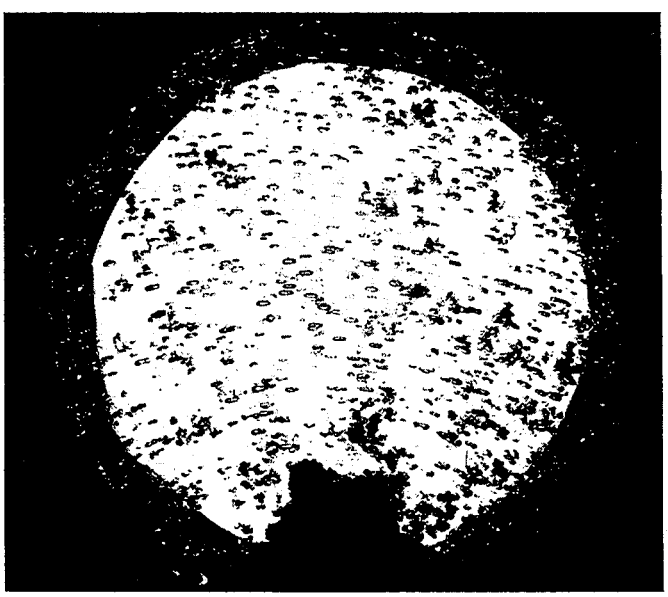


Fig. 6 (*e*). *12.5 milliseconds, weak radium source nearby:* vapor jet from the expansion orifice causes a sudden pressure wave in the chamber. It distorts the tracks and ends the sensitive time abruptly.

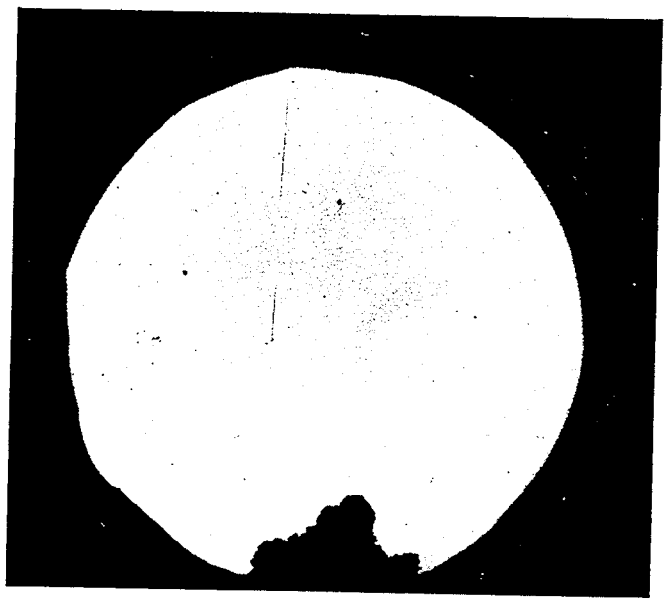


Fig. 6 (*b*). *12.5 milliseconds, no radiation:* boiling has progressed further and a jet of vapor shoots out of the expansion orifice at the bottom of the chamber.

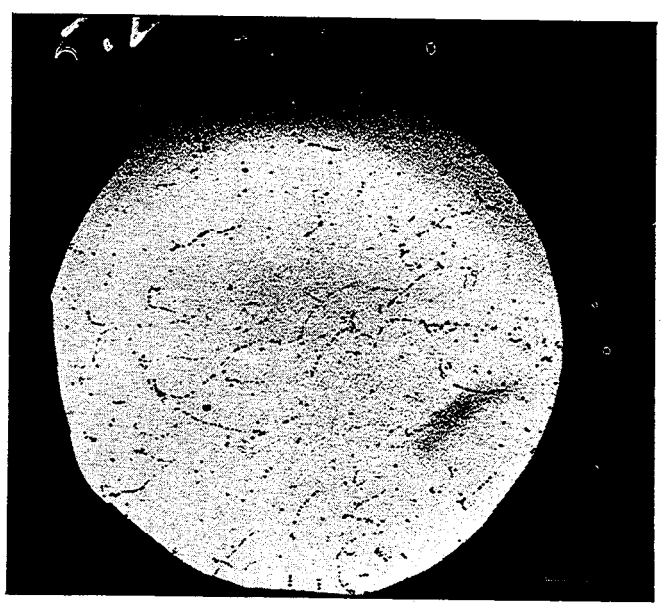


Fig. 6 (*c*). *5.4 milliseconds, radium source nearby:* some fine and some larger tracks are visible, the finer ones showing normal bubble density indicating that the chamber is at full sensitivity.

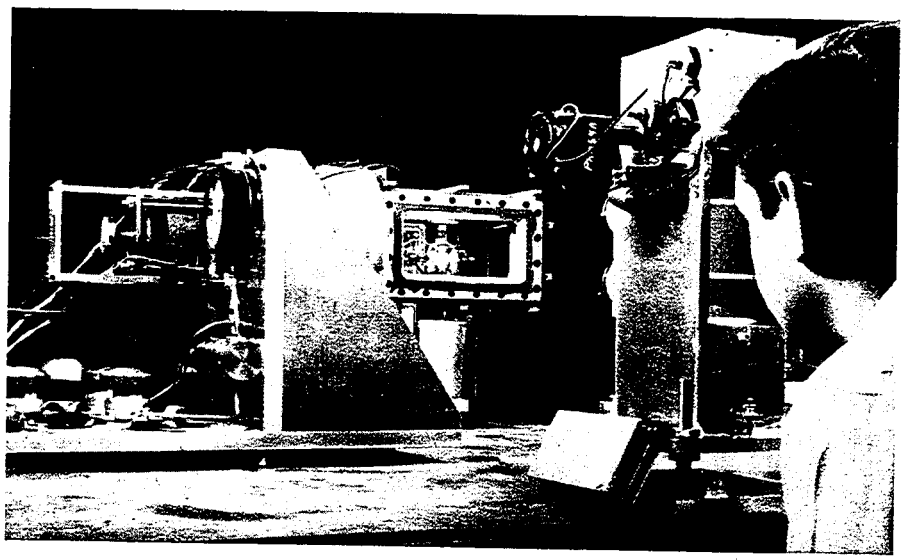


Fig. 7. A 15-cm propane bubble chamber showing stereoscopic camera, expansion system, and flashlamp. When in use, a small oven with windows fits around the chamber to maintain it at a temperature of 55°C to 60°C.

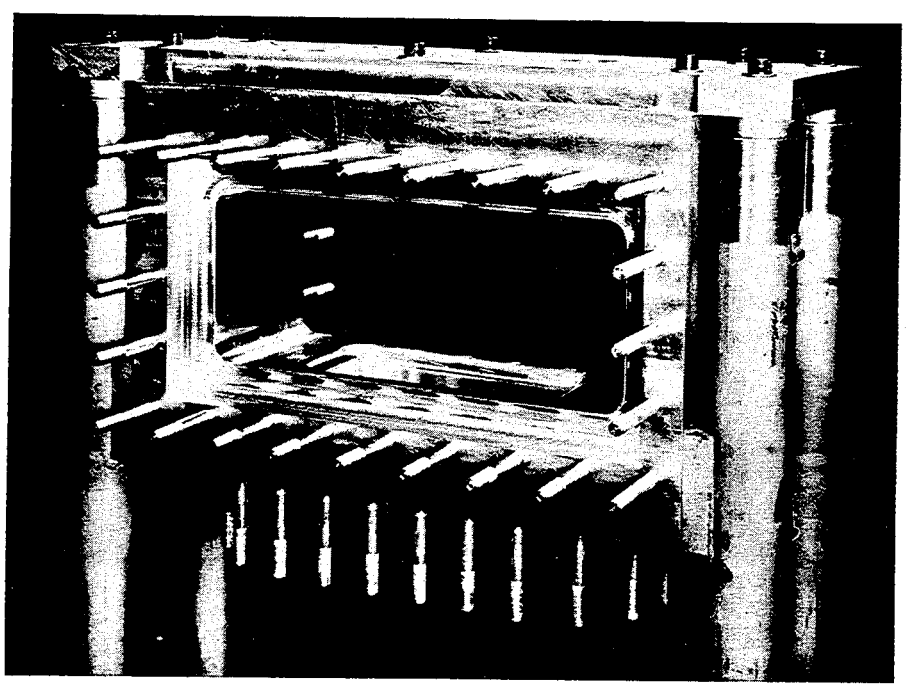


Fig. 8. The stripped aluminum body of a 30-cm propane bubble chamber.

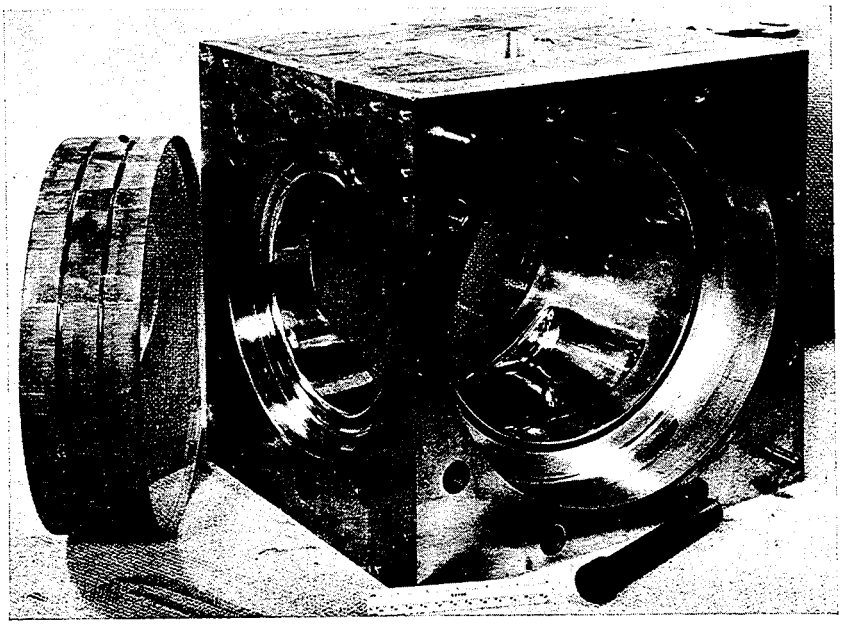


Fig. 9. The stripped aluminum body and test window for 21-liter, 30-cm xenon bubble chamber.

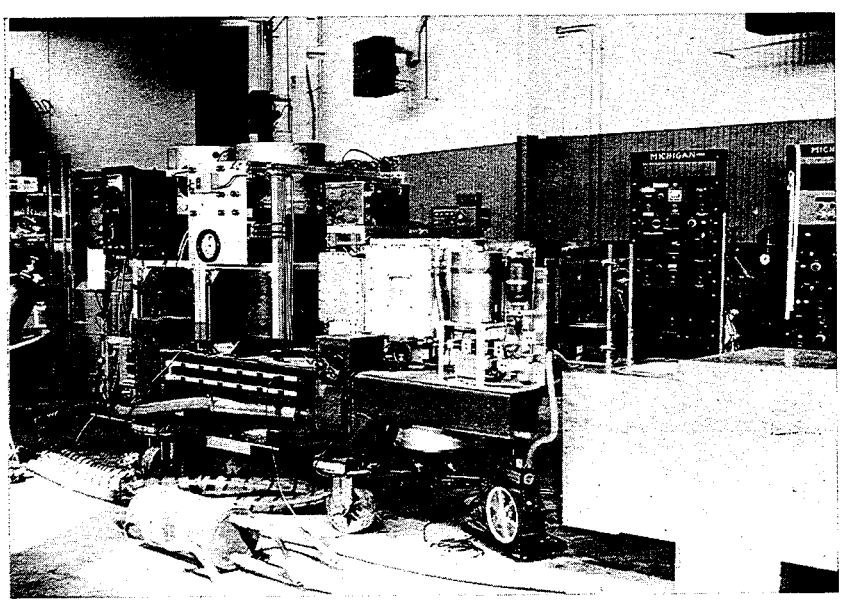


Fig. 10. Xenon bubble chamber assembled ready for use at the Bevatron, except for insulating refrigeration.

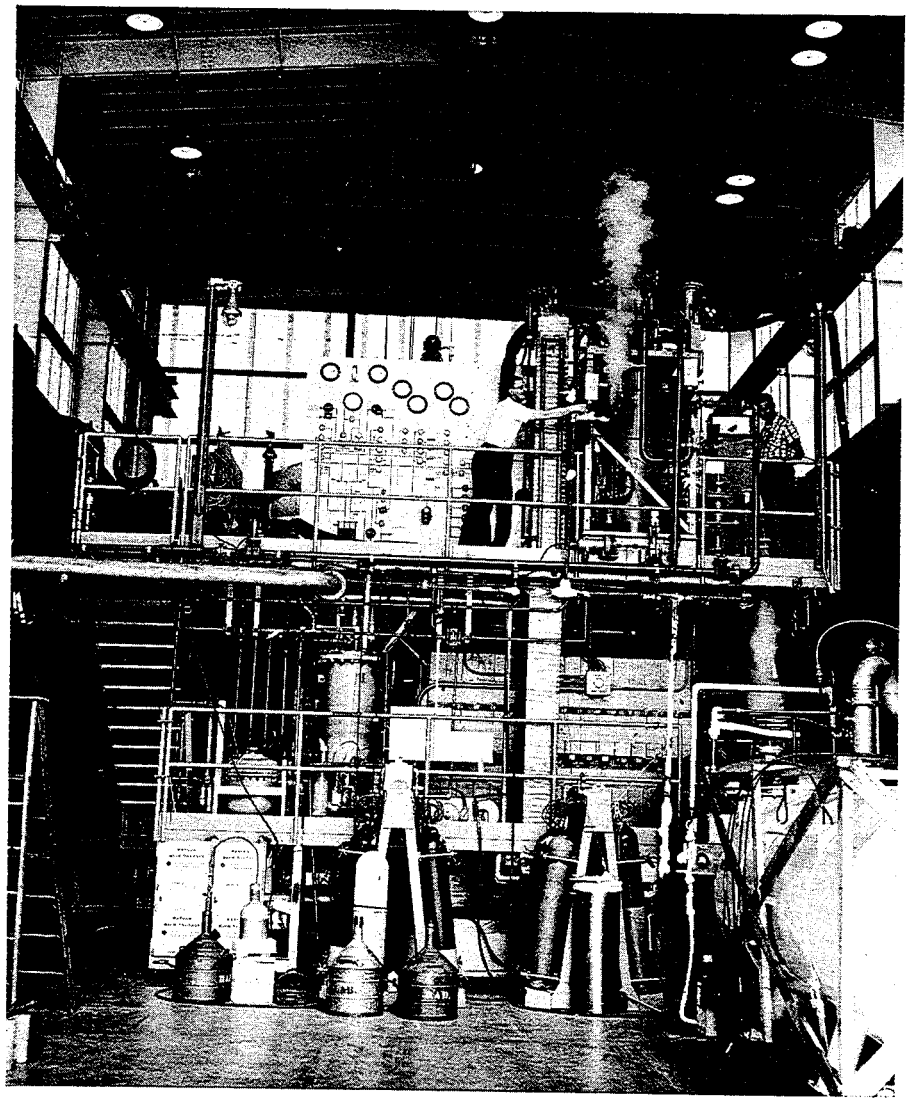


Fig. 11. Alvarez's 180-cm, 500-liter hydrogen bubble chamber assembled with magnet, ready for doing an experiment at Berkeley.

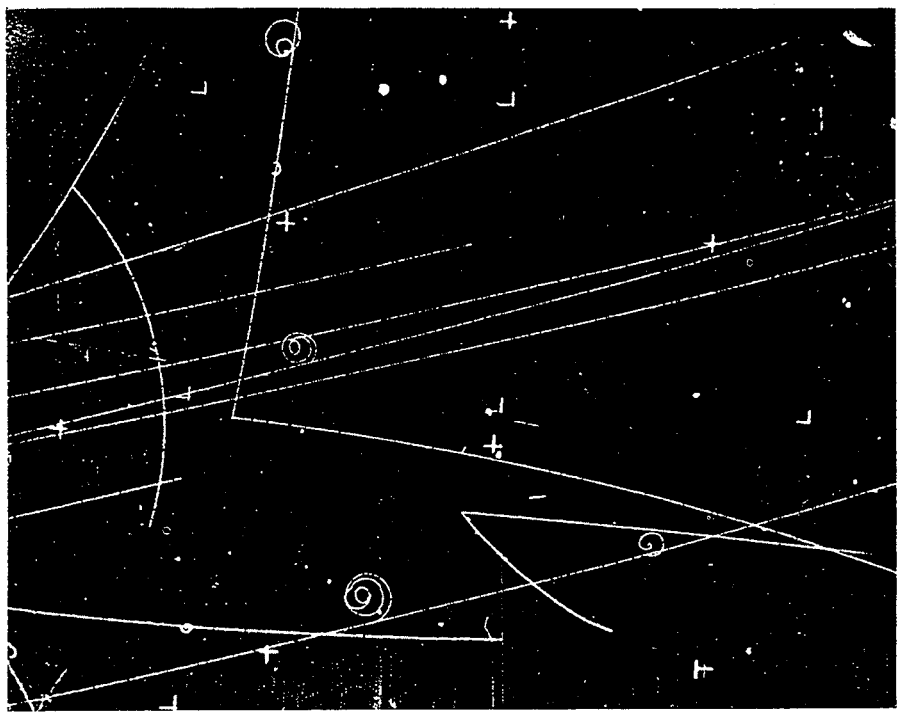


Fig. 12. Associated production, $\pi^- + p \rightarrow \Lambda^\circ + K^\circ$ at about 1 GeV with subsequent decays in Alvarez's hydrogen bubble chamber.

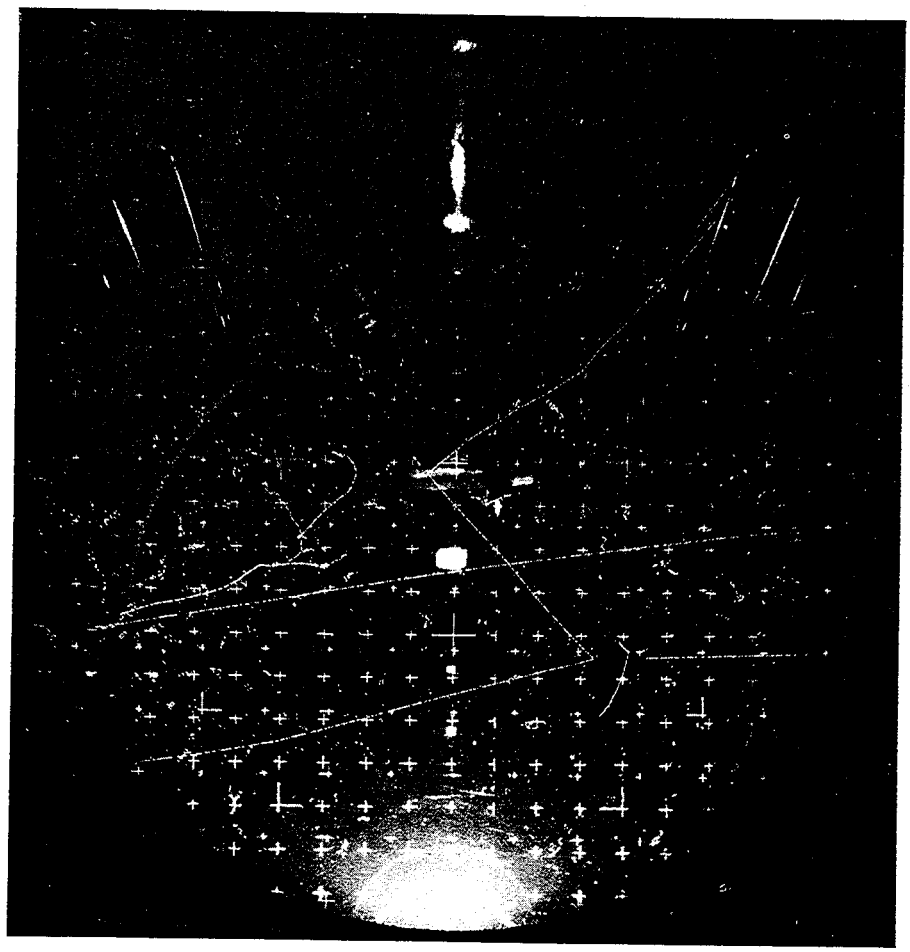


Fig. 13. Associated production, $\pi^- + p \rightarrow \Lambda^\circ + K^\circ$ at about 1 GeV seen in xenon bubble chamber with subsequent decays. The decay tracks of the Λ are very short because of the high density of the xenon.

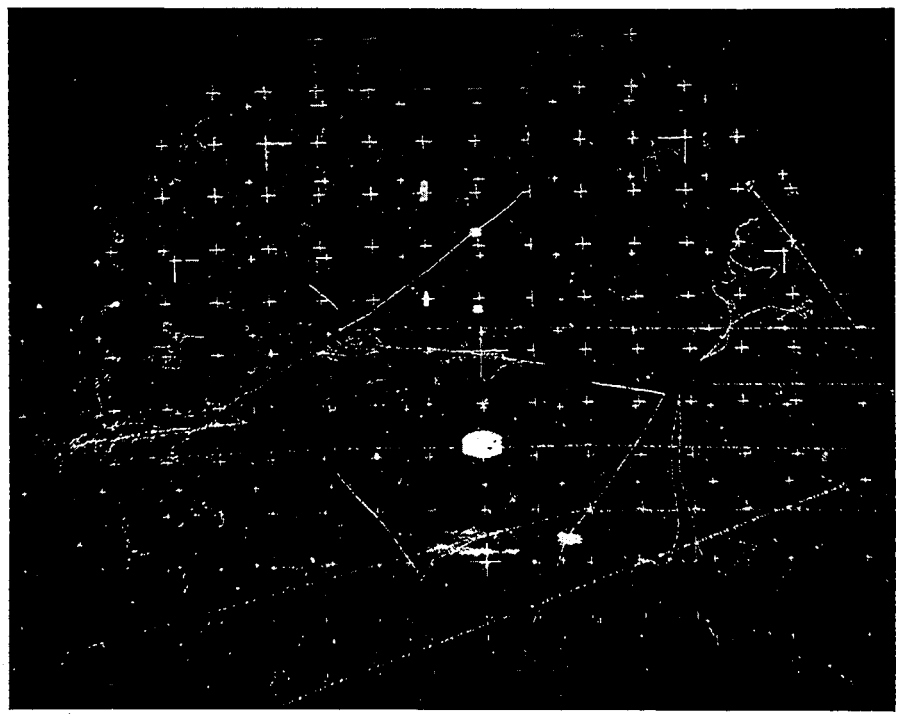


Fig. 14. The same process, $\pi^- + p \rightarrow \Lambda^\circ + K^\circ$ at about 1 GeV followed by $\Lambda^\circ \rightarrow p + \pi^-$ and $K^\circ \rightarrow \pi^\circ + \pi^\circ$ followed by $\pi^\circ \rightarrow 2\gamma$ in which all four gamma rays make electron showers. Complete processes like this involving gamma rays are efficiently detected in xenon because of its short radiation length.

chamber takes much longer than this, it is impossible to make expansions only when interesting events are detected by counter arrays because the bubble nuclei will have disappeared by the time the chamber has become sensitive. By random cycling and various ways of extending the length of the sensitive time, it may be possible to maintain a bubble chamber in the sensitive condition for 10% of the time. Beams of artificially produced particles can be timed to arrive at a bubble chamber just when it reaches full sensitivity. For use with particle accelerators the bubble chamber is ideal because it can be made to be sensitive only during the short time of beam arrival. General background ionization does not then complicate the pictures very much.

In its various forms the bubble chamber has solved many of the problems of obtaining large amounts of precise pictorial information on processes involving energetic elementary particles. At the same time it has created a new problem by providing each active laboratory with millions of photographs each year. These photographs require careful, intelligent inspection and measurement with coordinate accuracies of the order of 1 micron on the film. Measuring projection microscopes to do the last step have been constructed to be able to follow interesting tracks semi-automatically and punch the coordinate information onto cards which are later fed into high-speed digital electronic computers. From these computers come geometric, kinematic, and dynamical conclusions which identify known particles, measure energies, and identify known processes. Now we are faced with the problem of finding and measuring these events as fast as the bubble chamber and accelerator can produce them. Although we can contemplate staffs of 50 specialists engaged in this work at a single laboratory, or for a single large bubble chamber, a staff of 500 or 5,000 people seems impossible. Yet we can ask important questions in basic physics which would be answered if we had such a huge staff studying photographs. As a result, many bubble chamber physicists have turned their attention toward developing automatic pattern recognition, measuring, and computing machines. Some day it is dreamed that such a machine, armed with a memory filled with full knowledge of all known processes occurring in high-energy physics, will devour miles of film each day, duly noting the numbers, characteristics, and types of all the known processes it recognizes. Only when it cannot « understand » an event by searching its memory, will it ring a bell to call over a physicist who will try to understand the new process. A start has been made in this direction and in three to five years we expect to use the first machines capable of

Biography

Donald Arthur Glaser was born in Cleveland, Ohio, on September 21, 1926, the son of William J. Glaser, a businessman, and his wife Lena. He received his early education in the public schools of Cleveland Heights, Ohio, and took his B.Sc. degree in physics and mathematics at the Case Institute of Technology in 1946. His first original research is described in his bachelor's thesis and consists of an electron diffraction study of the properties of thin metallic films evaporated onto crystalline metal substrates.

After serving as a teacher of mathematics at the Case Institute of Technology during the spring of 1946, he began his graduate study at the California Institute of Technology in the autumn of the same year, finishing his Ph.D. work in the autumn of 1949, and receiving his degree in physics and mathematics officially in 1950. His doctoral thesis research was an experimental study of the momentum spectrum of high energy cosmic ray and mesons at sea level.

Glaser began his career of full-time teaching and research in the Physics Department of the University of Michigan in the autumn of 1949, being promoted to the rank of Professor in 1957. In 1959 he became Professor of Physics at the University of California, at Berkeley. His main research interest during this period was the elementary particles of physics, particularly the strange particles. He examined various experimental techniques that seemed useful in this research and constructed a number of diffusion cloud chambers and parallel-plate spark counters before finally beginning to develop the ideas that led to the invention of the bubble chamber in 1952. Since then he has worked on the development of various types of bubble chambers for experiments in high energy nuclear physics, besides carrying out experiments on elementary particles at the Cosmotron of the Brookhaven National Laboratory in New York and the Bevatron of the Lawrence Radiation Laboratory in California. These experiments gave information on the lifetimes, decay modes, and spins of the Λ° hyperon, K° meson and Σ° hyperon as well as differential cross-sections for the production of those particles by pions.

recognizing and interpreting a large class of known events. Armed with such formidable experimental means, the limitations of our imagination (and our ability to keep all these monstrous machines in operation) define the limits of the questions we may hope to answer about the experimental behavior of the elementary particles.

Since this lecture is intended to be a personal and historical description of research rather than a complete review article, the reader is referred to the following review sources for further and more detailed information.

D. A. Glaser, «The Bubble Chamber», *Handbuch der Physik*, S. Flügge (Ed.), Springer-Verlag, Berlin, 1958, Vol. 45, p. 314.
H. Slätis, « Survey Article on Bubble Chambers », *Nucl. Instr. Methods*, 5 (1959) 1.
D. V. Bugg, «The Bubble Chamber», *Progress in Nuclear Physics*, O. R. Frisch (Ed.), Pergamon Press, London, 1959, Vol. 7, p. 1.

Other experiments yielded information on pion-proton scattering, parity violation in non-leptonic hyperon decay, and the branching ratios in positive K meson decay.

All these experiments and technical developments of the past six years were carried out in collaboration with a number of his thesis students and colleagues at the University of Michigan and the University of California at Berkeley, where he worked from 1959. Among his associates in research were J. Brown, H. Bryant, R. Burnstein, J. Cronin, C. Graves, R. Hartung, J. Kadyk, D. Meyer, M. Perl, D. Rahm, B. Roe, L. Roellig, D. Sinclair, G. Trilling, J. van der Velde, J. van Putten and T. Zipf.

These researches were supported originally by the University of Michigan and later by the National Science Foundation of the United States and the United States Atomic Energy Commission.

Glaser has received many honours for his work, among which can be mentioned the Henry Russell Award of the University of Michigan, 1953, for distinction and promise in teaching and research; the Charles Vernon Boys Prize of the Physical Society, London, in 1958, for distinction in experimental physics; the American Physical Society Prize (sponsored by the Hughes Aircraft Company) for his contributions to experimental physics in 1959; and the award, in the same year, of the honorary degree of Doctor of Science by the Case Institute of Technology.

1960, the year in which he was awarded the Nobel Prize for Physics, also saw Professor Glaser's marriage to Miss Ruth Bonnie Thompson.

Physics 1961

ROBERT HOFSTADTER

«for his pioneering studies of electron scattering in atomic nuclei and for his thereby achieved discoveries concerning the structure of the nucleons»

RUDOLF LUDWIG MÖSSBAUER

«for his researches concerning the resonance absorption of γ-radiation and his discovery in this connection of the effect which bears his name»

Physics 1961

Presentation Speech by Professor I. Waller, member of the Swedish Academy of Sciences

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

Since Rutherford's discovery of the atomic nucleus fifty years ago, one of the most fundamental problems in physics has been to investigate how it is constituted. The ideas on this question could be firmly founded when, shortly after 1930, a neutral particle called the neutron was discovered which has almost the same mass as the hydrogen nucleus i.e. the proton. A theory for the atomic nuclei was proposed according to which they are composed of protons and neutrons which are together called the nucleons. A few years later, Yukawa gave a theory of the forces which keep the nucleons together. It could according to this theory be expected that the nucleons have themselves a complicated inner structure.

Professor Robert Hofstadter has developed a new experimental method for the investigation of the inner structure of the composite atomic nuclei and also of the single nucleons. His method is to bombard the atomic nuclei with electrons of very high energy. The electrons can penetrate the atomic nuclei and are then deviated by the strong electric and magnetic forces inside the nuclei. By separating the scattered electrons of different energies in magnetic spectrometers and by measuring afterwards the number of electrons which have been deviated to each particular direction, Hofstadter has succeeded in obtaining detailed knowledge of the distribution of the electric charge in the nuclei. For the nucleons, important results have also been found concerning the distribution of their magnetic moments.

The experimental method used by Hofstadter is connected with the principles of the ordinary electron microscope. Here the possibilities to observe details are increased by raising the voltage which accelerates the electrons. As the dimensions of the atomic nuclei are of the order of a ten-billionth of a centimeter, Hofstadter had in order to find their structure to bombard them with electrons of a very high energy. The highest energy used was equivalent to an accelerating voltage of nearly one billion volts. When Hofstadter in 1950 began his work at Stanford University a linear accelerator had already been constructed there and was later supplemented to give the

electrons this energy. Hofstadter has built a complicated experimental installation in order to make possible the measurements of the scattering with the accuracy required. He has in a very skilful manner managed to achieve great precision in spite of the very large dimensions of the apparatus.

Hofstadter's results have opened fundamentally new aspects on the properties of the nucleons. His investigations form a pioneering work which in the last years has been beautifully confirmed by similar experiments at Cornell University. It must be expected that investigations of this kind will be made at other institutes also. Electron accelerators which are expected to come into operation in some years will probably increase further our knowledge in this field.

Professor Rudolf Mössbauer's investigations concern the emission and absorption of gamma radiation by the atomic nuclei. This radiation is of the same kind as the light and the radio waves. It is well known that incoming radio waves can be received only if the receiver is tuned to the same frequency as the sender. Resonance is then taking place. It has since long been tried to observe the corresponding phenomenon for nuclei, where it is called « resonance absorption ». The method was to let gamma radiation from some kind of nuclei act upon other nuclei of exactly the same kind. There is however a certain difficulty connected with this experiment. The gamma radiation can be considered as made up of particles. When emitting a gamma particle the atom receives a recoil whereby the energy and therefore also the frequency of the gamma radiation is decreased. The same phenomenon occurs when the gamma particle is absorbed in the receiving nucleus. The resonance will be completely destroyed if the frequency change is not compensated for, as had been done already before Mössbauer's work. Mössbauer discovered experimentally and showed also theoretically that for atoms bound in a solid, an appreciable part of the radiation can be emitted without frequency change whereby the resonance absorption can be studied directly. This discovery was published by Mössbauer in 1958. Because of the very small width of the gamma lines the resonance is very sharp and can, as Mössbauer found, be influenced and finally inhibited by the Doppler effect if the source or the absorber for the gamma radiation is moved. The velocities required depend upon the sharpness of the gamma line and can be as small as some millimeters per hour.

Mössbauer's discovery has been received with considerable interest. Research on the Mössbauer effect has been started at a great number of places. It has thereby been possible to verify in the laboratory, fundamental con-

sequences of Einstein's theory of relativity. Other important applications depend on the separation and displacement of nuclear energy levels which occur in solids because of the influence of the surroundings. Many phenomena of this kind can in spite of their smallness be studied by the Mössbauer effect. It has been possible in this way to get most important information on the properties of solids.

Mössbauer made his discovery when he investigated the resonance absorption on the suggestion of Professor Maier-Leibnitz in München. He found then some unexpected results which he investigated systematically and was thereby led to his discovery.

Professor Hofstadter. You have in your pioneering investigations on the atomic nuclei and the single nucleons, revealed features of their structures which are fundamentally important for the understanding of these almost inconceivably small systems. Your work is characterized by a precision which has scarcely been attained before in high-energy physics. You have achieved this precision by improving unrelentingly your methods and equipment in the course of time. Your results have quite recently stimulated the discovery of new particles which seem to be essential for the understanding of the forces acting in the atomic nuclei.

Professor Mössbauer. While doing research for your doctor's thesis you have discovered an unexpected effect which now bears your name. You have explained this effect experimentally and theoretically, and thereby created a device which is of fundamental importance in numerous realms of physics, and which is nowadays being investigated and put to use in a large number of physical laboratories. By your discovery it has become possible to examine precisely, numerous important phenomena formerly beyond or at the limit of attainable accuracy of measurement.

Professor Hofstadter and Professor Mössbauer. May I congratulate you on behalf of the Academy on your important work and ask you to receive the Nobel Prize from the hands of His Majesty the King.

ROBERT HOFSTADTER

The electron-scattering method and its application to the structure of nuclei and nucleons

Nobel Lecture, December 11, 1961

I am very conscious of the high honor that has been conferred on me and I wish to thank the Swedish Academy of Sciences sincerely for this recognition. It is a privilege and a pleasure to review the work which has brought me here and which concerns a very old and interesting problem.

Over a period of time lasting at least two thousand years, Man has puzzled over and sought an understanding of the composition of matter. It is no wonder that his interest has been aroused in this deep question because all objects he experiences, including even his own body, are in a most basic sense special configurations of matter. The history of physics shows that whenever experimental techniques advance to an extent that matter, as then known, can be analyzed by reliable and proved methods into its « elemental » parts, newer and more powerful studies subsequently show that the « elementary particles » have a structure themselves. Indeed this structure may be quite complex, so that the elegant idea of elementarity must be abandoned. This observation provides the theme of our lecture.

In recent times the structure of matter has been shown to arise from various combinations of the « atoms » of the Periodic System. The picture of the now-familiar atom was first sketched by Rutherford, Bohr, Pauli, etc., and later developed in great detail by many of their colleagues. The efforts of these scientists have led to an understanding of the cloud of electrons which surrounds the dense center of the atoms, the so-called nucleus. In the nucleus practically all the mass of the atom resides in an extremely concentrated form. The nucleus itself was an invention of the aforementioned physicists and in the year 1919 the first vague ideas concerning the sizes of nuclei were worked out. By studying the deviations from Coulomb scattering of alpha particles Rutherford showed that a nuclear radius was of the order of 10^5 times smaller than an atomic radius. Subsequently other investigators demonstrated by means of studies of α-particle radioactivity, neutron capture cross sections, and comparisons of the energy of decay of mirror nuclei that consistent values of nuclear size parameters could be measured.

All useful methods showed that if a nucleus could be represented by a model of a uniformly charged sphere the radius (R) of the sphere would be given by the relation

$$R \cong 1.40 \cdot 10^{-13} A^{\frac{1}{3}} \text{ cm} \tag{1}$$

where A is the mass number of the nucleus.

This is the point from which the present studies began. Although much of what we wish to say will concern nucleon structure (nucleon = proton or neutron) the method of investigation we have employed had its origins in the study of larger nuclei. Consequently a historical approach beginning with the larger nuclei seems not only natural but also may be didactically sound. We shall therefore review briefly the method used in studying nuclear sizes and shall at the same time give some of the results, which may not be without interest themselves.

We have used the method of high-energy electron scattering. In essence the method is similar to the Rutherford scattering technique, but in the case of electrons it is presently believed that only a « simple » and well-understood interaction – the electromagnetic or Coulomb interaction – is involved between the incident electron and the nucleus investigated. Under these conditions quantum electrodynamics and Dirac theory teach us how to calculate a differential elastic scattering cross section. It can be shown that the differential cross section corresponding to a beam of electrons scattering against a point nucleus of small charge Ze, lacking spin and magnetic moment, is calculable by the Born approximation and takes a form:

$$\frac{d\sigma}{d\Omega} = \left(\frac{Ze^2}{2E}\right)^2 \frac{\cos^2 \theta/2}{\sin^4 \theta/2} \frac{1}{1 + \frac{2E}{Mc^2} \sin^2 \theta/2} = \sigma_{NS} \tag{2}$$

in the laboratory system of coordinates[1]. This is the « Mott » scattering cross section, where E is the incident energy, θ the scattering angle and M the mass of the struck nucleus. Other symbols in Eq. 2 have their usual meanings. If a nucleus has a finite size, and is thus not merely a point, the scattering cross section is decreased below the value of the scattering from a point. The decrease can be described in terms of a factor, represented by F, which is called the « form factor » or « structure factor ». Thus, in Born approximation,

$$\frac{d\sigma}{d\Omega} = \sigma_{NS} F^2(q) \tag{3}$$

and this is the elastic scattering cross section for a finite nucleus[2]. Here q is the momentum–energy transfer, defined by the relation

$$q = \frac{(2E/\hbar c)\sin\theta/2}{\sqrt{1+(2E/Mc^2)\sin^2\theta/2}} \tag{4}$$

The parameter q is relativistically invariant and is a very important quantity in electron-scattering studies. The form factor, F, takes account of the interference between scattered wavelets arising from different parts of the same, finite, nucleus and therefore is responsible for diffraction effects observed in the angular distribution. The quantity F is actually given by

$$F = \frac{4\pi}{q}\int_0^\infty \varrho(r)(\sin qr)\,r\,dr \tag{5}$$

in the event that the nucleus exhibits spherical symmetry. The quantity $\varrho(r)$ is the electric charge density function, in which r represents the distance from the center of the nucleus to the volume element where ϱ is measured. A mathematical inversion of Eq. 5 allows one to deduce the form of $\varrho(r)$ if $F(q)$ is known over a large range of values of q.

Of course, since we used the Born approximation and therefore specified small values of the atomic number, the above description of the basic formulae of the electron-scattering process is only an approximate one. More exact methods of finding the scattering cross section have been developed by many authors[3]. These calculations of more precise types employ the « phase-shift » methods and are applicable to heavy nuclei as well as light ones. The qualitative physical ideas involved in the determination of nuclear structure can be adequately described by the Born approximation method (Eq. 3). Nevertheless, quantitative results definitely require the more elaborate phase-shift methods and simple, and in this case, closed formulae cannot be given to describe the scattering cross section.

Early electron-scattering experiments were carried out at the University of Illinois in 1951[4] at an incident electron energy of about 15.7 MeV. Such experiments showed that nuclear radii obeyed an approximate relationship of the type given in Eq. 1. However, few details of nuclear shape or size could be discerned because the energy of the electrons was relatively low and the corresponding De Broglie wavelength of the electrons was larger than a typical size of the nucleus. In 1953 higher energy electrons became

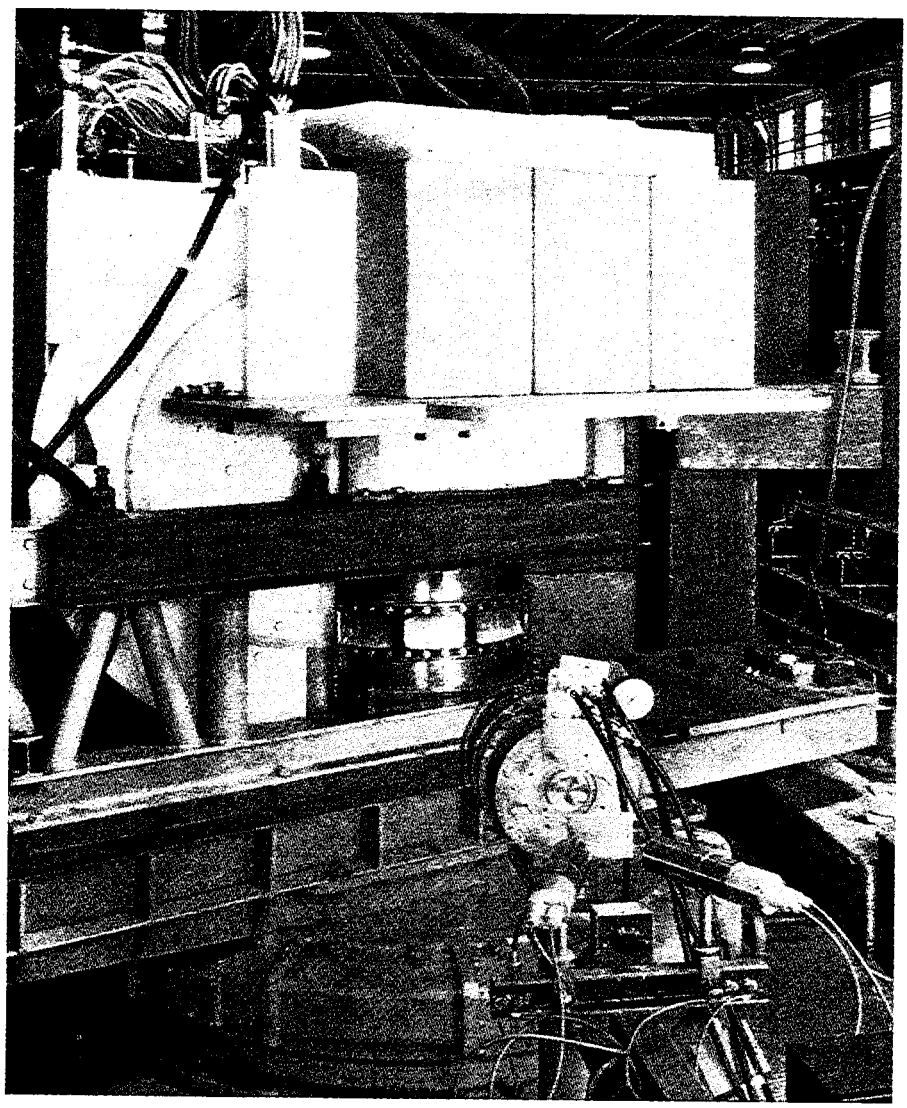


Fig. 1. The first electron-scattering apparatus built at Stanford University. The semicircular 190-MeV spectrometer is shown at the left on its gun-mount support. The upper platform carries lead and paraffin shielding that encloses the Čerenkov counter. The brass scattering chamber is shown below with the thin window encircling it. Early forms of electron monitors appear in the foreground. The spectrometer itself is about four feet high.

available at Stanford University and at the University of Michigan and experiments on various nuclei were carried out[5]. Phase-shift interpretations of the Stanford experiments[6] showed that the rule expressed in Eq. 1 was approximately true, but that in reality the nuclear charge density distribution could not be described in terms of a single size parameter R. If one attempted to do so, only at the expense of an inferior fit between experiment and theory, the resulting R would have to be made 20% smaller than the value of the radius in Eq. 1. Mu-mesonic atom studies[7] showed, a bit earlier, that a similar conclusion was required for a one-parameter description of the size of the nucleus. Two parameters could not be determined from the mu-mesonic atom investigations.

Fig. 1 shows a photograph of the first high-energy electron-scattering equipment. This apparatus gave the above results and was employed up to an energy of about 190 MeV. An obsolete naval-gun mount was used as the rotating platform for the heavy equipment, weighing about 5 tons. The type of geometry employed in a modern electron-scattering experimental area is shown in Fig. 2. A photograph of the corresponding magnetic spectrometers and associated equipment is shown in Fig. 3. A larger form of gun mount is used again to carry the spectrometers, whose total weight is approximately

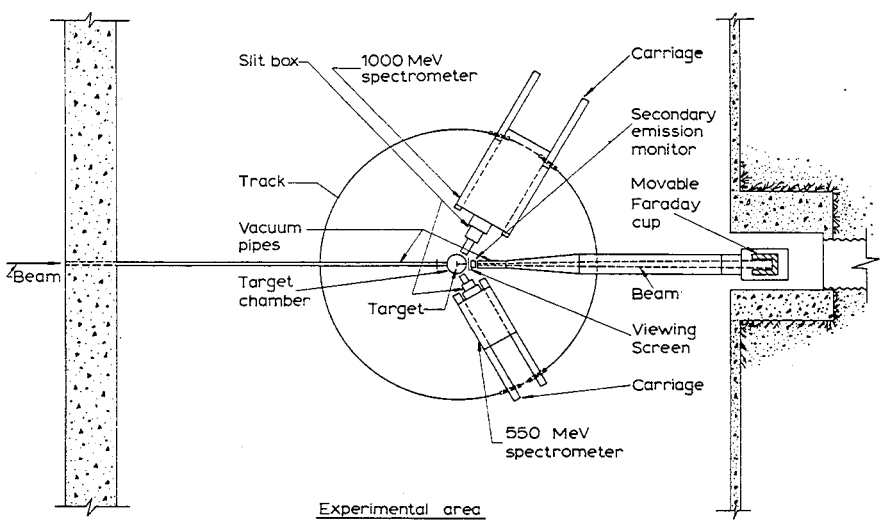


Fig. 2. This figure shows a schematic diagram of a modern electron-scattering experimental area. The track on which the spectrometers roll has an approximate radius of 13.5 feet.

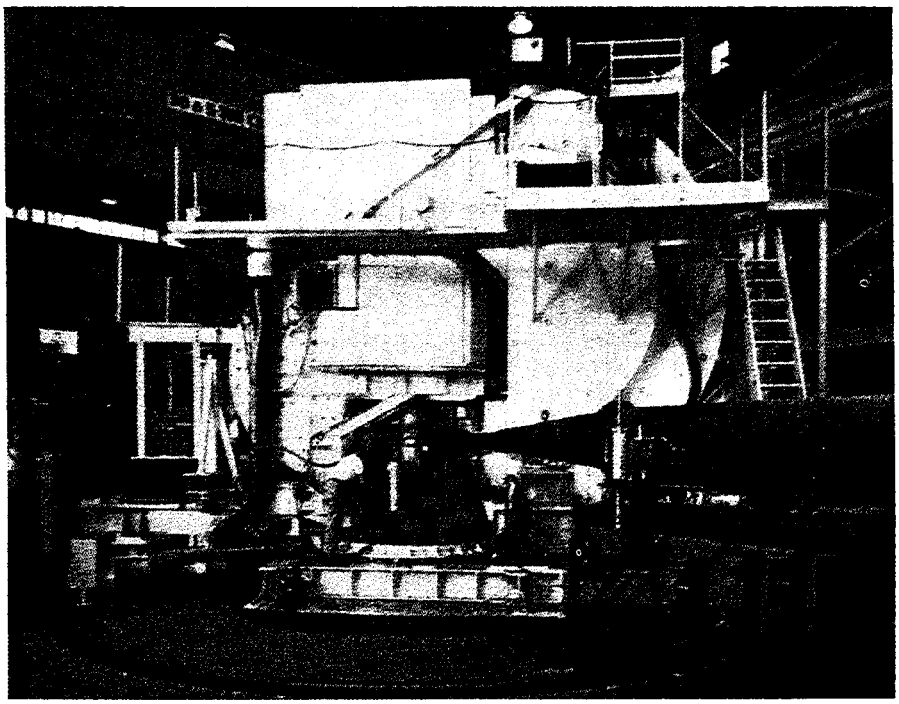


Fig. 3. A recent photograph of the double-spectrometer system is shown in this figure. The shield of the smaller spectrometer can be removed easily with the aid of an auxiliary stand, not shown in the photograph. The long tube in the foreground is the vacuum pipe leading to the Faraday cup, which is not visible in the photograph.

250 tons. Each of the two magnetic spectrometers in this apparatus is similar to the well-known Siegbahn double-focussing instrument. The two spectrometers may be used in coincidence experiments as well as « in parallel ». The massive equipment of Fig. 3 can bend and focus 1.0-BeV electrons and is required in order to resolve the elastic-scattering process from the many types of inelastic-scattering processes occurring in electron–nucleus collisions. An example of the resolution obtained in early experiments is shown in Fig. 4 in the case of a carbon target[8]. When an angular distribution in carbon is measured one may observe, e.g. in Fig. 5, the position of a diffraction minimum. The value of the angle at this minimum gives immediately an indication of the nuclear size if one employs results similar to Eqs. 2–5, when modified appropriately in terms of the phase-shift method. The solid line in the figure shows the result of a theoretical calculation of the scattering cross

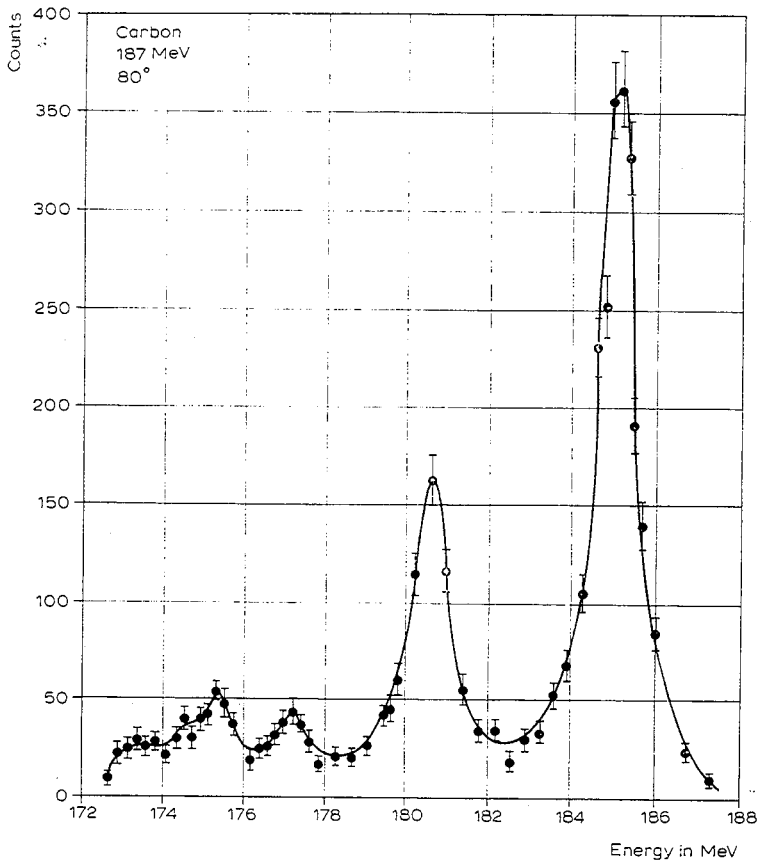


Fig. 4. This figure[8] shows the elastic-scattering peak from carbon at an abscissa near 185 MeV, and the inelastic-scattering peaks from the excited states of ^{12}C. The peak near 180.7 MeV is associated with the 4.43-MeV level.

section. From the theoretical calculation one may deduce the charge density distribution, which may be seen in Fig. 8. It is clear that a study of the inelastic-scattering peaks corresponding to the excited states of ^{12}C (or other nuclei) can be studied by the electron-scattering method. In fact, Fig. 5 shows also the angular dependence of the scattering of the 4.43-MeV level in ^{12}C. The subject of inelastic level scattering is not relevant to our present topic and we shall not pursue this matter any further in this lecture.

One last example is shown in the case of the nucleus of the gold atom. The elastic electron scattering was studied at the four different energies shown in Fig. 6. The solid lines again show the results of theoretical calculations from

which the charge density distribution, ϱ, can be obtained. This charge distribution is shown in Fig. 8.

The electron-scattering method was employed in the manner we have described and resulted in the determination of two-parameter descriptions of nuclear charge density distributions. Studies of the charge density dis-

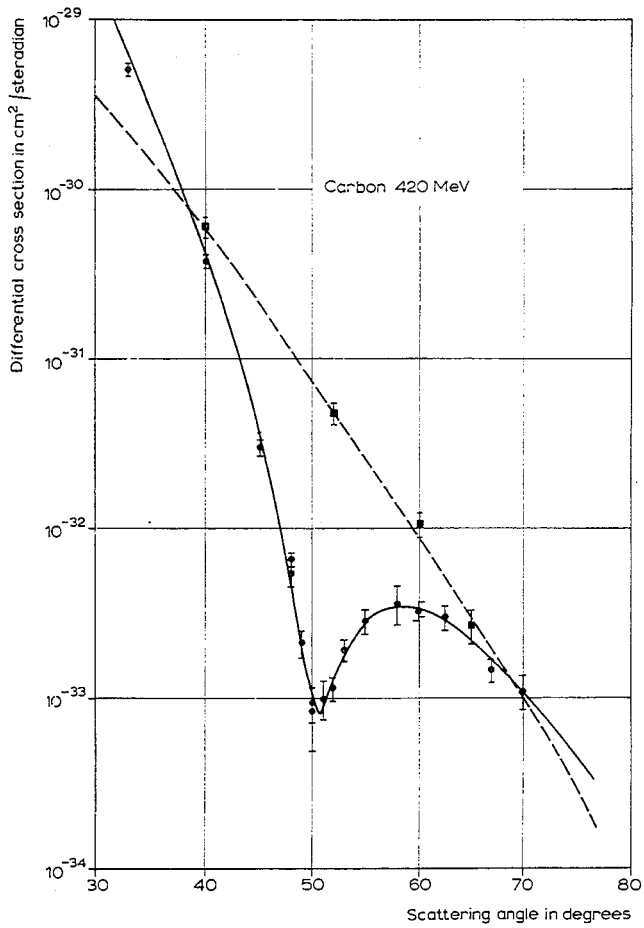


Fig. 5. This figure shows the elastic and inelastic curves corresponding to the scattering of 420-MeV electrons by ^{12}C. The *solid circles*, representing experimental points, show the elastic-scattering behavior while the *solid squares* show the inelastic-scattering curve for the 4.43-MeV level in carbon. The *solid line* through the elastic data shows the type of fit that can be calculated by phase-shift theory for the model of carbon shown in Fig. 8.

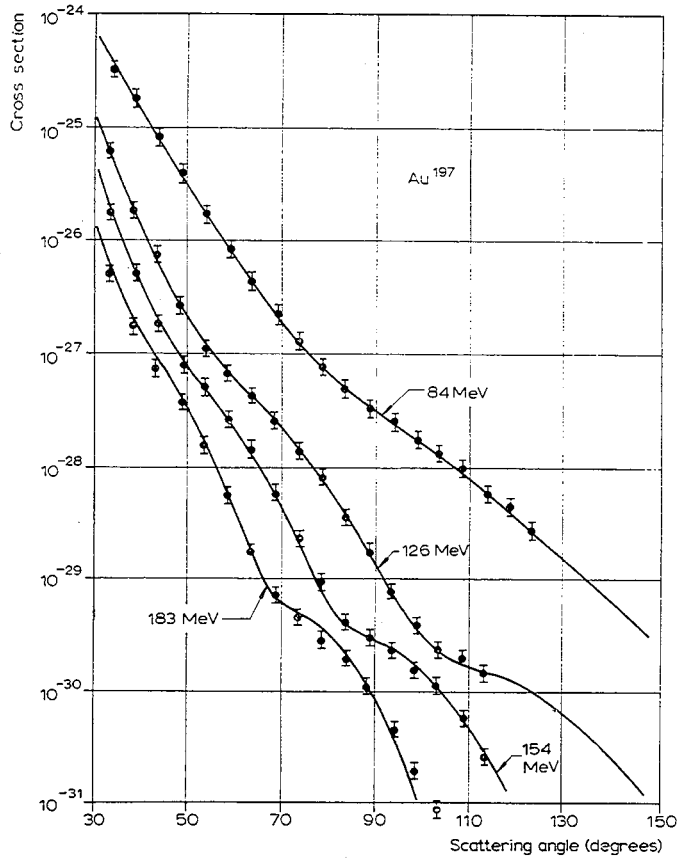


Fig. 6. The *points* represent experimental data observed by scattering electrons of the appropriate incident energies from gold nuclei[9]. The *solid lines* are calculated angular distributions for a model of the gold nucleus similar to that shown in Fig. 8.

tributions in various nuclei culminated in the evolution of a simple scheme of construction of most spherical nuclei[9]. Such nuclei could be represented by a charge density function of the type shown in Fig. 7. The exact shape of this density function is not of overriding importance; rather the distance (c) from the center of the nucleus to the 50 per cent point, and the interval (t) between the 90 per cent and 10 per cent ordinates, are the two important parameters that determine the behavior of the scattering cross sections. A trapezoidal distribution with the same values of the two parameters would also suffice to describe the experimental results in the medium and heavier nuclei when the fitting procedure is limited by the accuracy obtained in the

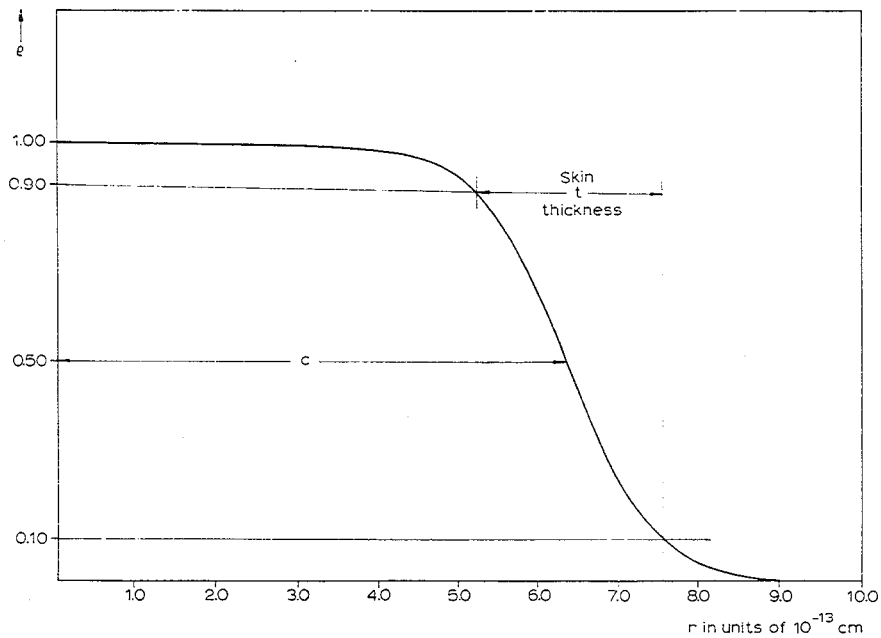


Fig. 7. The shape and parameters which describe an approximate model of the gold nucleus. This type is called the Fermi model[9].

experiments. Higher accuracy can probably distinguish between these possibilities but such studies are only beginning now.

The results of many of the above experiments covered a large range of nuclei and demonstrated[9] that two simple rules can be used to summarize the scheme of construction of spherical nuclei, viz.:

$$\begin{aligned} c &= (1.07 \pm 0.02) \cdot 10^{-13} A^{\frac{1}{3}} \text{ cm} \\ t &= (2.4 \pm 0.3) \cdot 10^{-13} \text{ cm} = \text{constant} \end{aligned} \qquad (6)$$

The first equation gives the principal parameter governing the size of a nucleus and describes the behavior with increasing A of a kind of «mean» nuclear radius. The second equation states that the nuclear skin thickness is constant. The second rule implies that there is some property of nuclear matter that causes the outer nuclear regions to develop an essentially constant surface thickness. The two rules together are responsible for the approximate constancy of the central charge density of nuclei. The latter property is il-

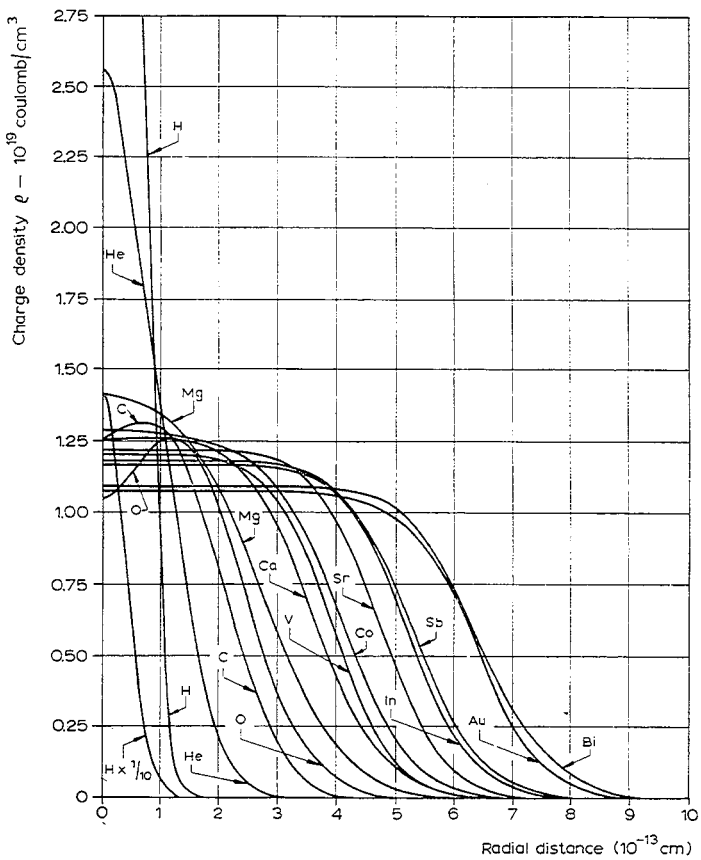


Fig. 8. This figure gives a summary of the approximate charge density distributions found for various nuclei studied by electron-scattering methods. The central densities are the least well determined positions of the curves. Note, however, the large disparity between the *average* central densities of the proton and all other nuclei. The alpha particle (^{4}He) is also a unique case and exhibits a much larger central density than all heavier nuclei.

lustrated in Fig. 8, where a summary of the charge distributions found by the electron-scattering method is presented for various nuclei. Except for the extremely light nuclei of hydrogen and helium the constancy of the central nuclear density is clearly represented in the figure.

The results obtained with heavier nuclei indicated that the electron-scattering method could also be applied to the very light nuclei and even to the proton itself. Accordingly, in early 1954 experiments were initiated on hy-

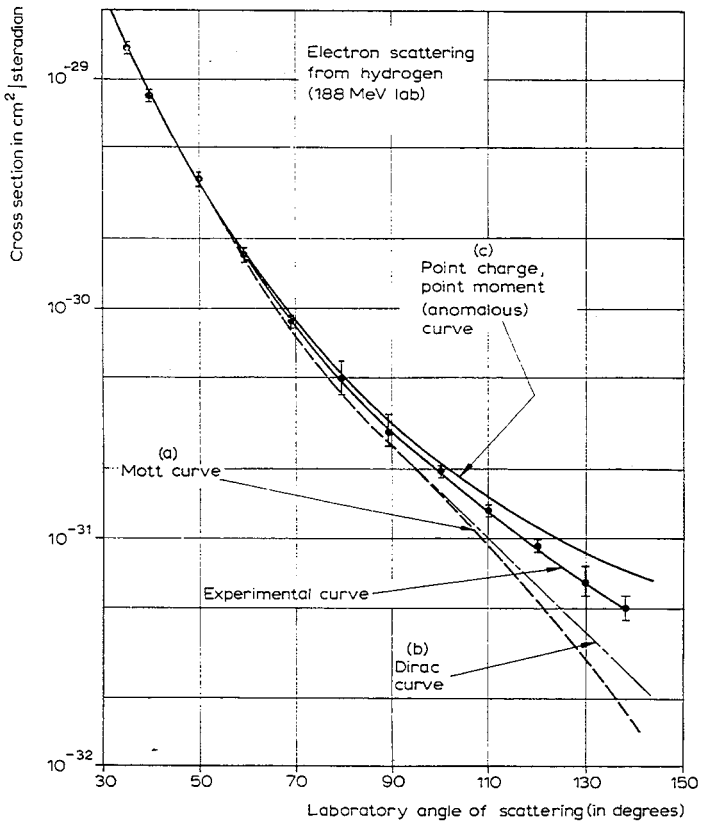


Fig. 9. Electron scattering from the proton at an incident energy of 188 MeV. *Curve* (*a*) shows the theoretical Mott curve for a spinless point proton. *Curve* (*b*) shows the theoretical curve for a point proton with a Dirac magnetic moment alone. *Curve* (*c*) shows the theoretical behavior of a point proton having the anomalous Pauli contribution in addition to the Dirac value of the magnetic moment. The deviation of the experimental curve from the Curve (c) represents the effect of form factors for the proton and indicates structure within the proton. The best fit in this figure indicates an rms radius close to $0.7 \cdot 10^{-13}$ cm.

drogen and helium. The first targets employed high-pressure, thin-wall, gas chambers and were designed by the late Miss Eva Wiener. In the latter part of 1954 it was first realized that the experiments on hydrogen demonstrated that the proton was an object of finite size and not merely a point object. In fact, the size was found to be surprisingly large[10] and could be described in terms of a root-mean-square radius of value $(0.74 \pm 0.24) \cdot 10^{-13}$ cm. It is an interesting fact that more recent determinations of the rms proton charge

radius appear to converge on a value of $(0.79 \pm 0.08) \cdot 10^{-13}$ cm. Fig. 9 shows the first evidence of finite size in the proton. The figure has been drawn from Ref. 10. The first experiments leading to the above conclusions were carried out at relatively low energies ($\sim$ 190 MeV).

Now the proton is known to have a spin and a magnetic moment. The magnetic moment will affect the scattering behavior appreciably at values of q (Eq. 4) in the range equal to or larger than about 0.2 Mc, where M is the mass of a nucleon. The magnetic type of scattering causes a leveling off in the decrease of the elastic cross section as a function of the scattering angle at high energies of the incident electrons. As we may see in Fig. 9, the experimental data fell below the expected theoretical curve for a proton possessing a point charge and a point magnetic moment. This behavior can be understood in terms of the theoretical scattering law developed by M. Rosenbluth[11] in 1950. This law described the composite effect of charge and magnetic moment scattering and is given by:

$$\frac{d\sigma}{d\Omega} = \sigma_{NS} \left\{ F_1^2 + \frac{\hbar^2 q^2}{4 M^2 c^2} \left[2 (F_1 + K F_2)^2 \tan^2 \theta/2 + K^2 F_2^2 \right] \right\} \quad (7)$$

where σ_{NS} is taken from Eq. 2 with $Z = 1$. In the Rosenbluth equation the quantity $F_1(q)$ is the Dirac form factor, representing the proton's charge and its associated Dirac magnetic moment. The quantity $F_2(q)$ is the Pauli form factor and represents the anomalous magnetic moment of the proton. K in the above equation indicates the static value (1.79) of the anomalous magnetic moment in nuclear magnetons.

Although one may speak qualitatively of size and shape factors of the proton in the low-energy limit it is more consistent and more desirable, from a quantitative point of view, to discuss only the two phenomenological form factors $F_1(q)$ and $F_2(q)$. Actually all the electromagnetic structure of the proton is, in principle, described by the behavior of these quantities as functions of q. Note that for the proton, $F_1(0) = F_2(0) = 1.00$. Meson theory should be able to make definite assertions about F_1 and F_2 starting from the above values. In our subsequent discussion we shall concentrate on determining the two phenomenological quantities (F_1, F_2) from the experimental data so that the form factors can be compared with theory. The experimental determinations of the form factors can be accomplished, for example, by using the method of intersecting ellipses[12] or by other equivalent methods based on the relativistic idea that each F is a function *only* of q and not of E or θ separately.

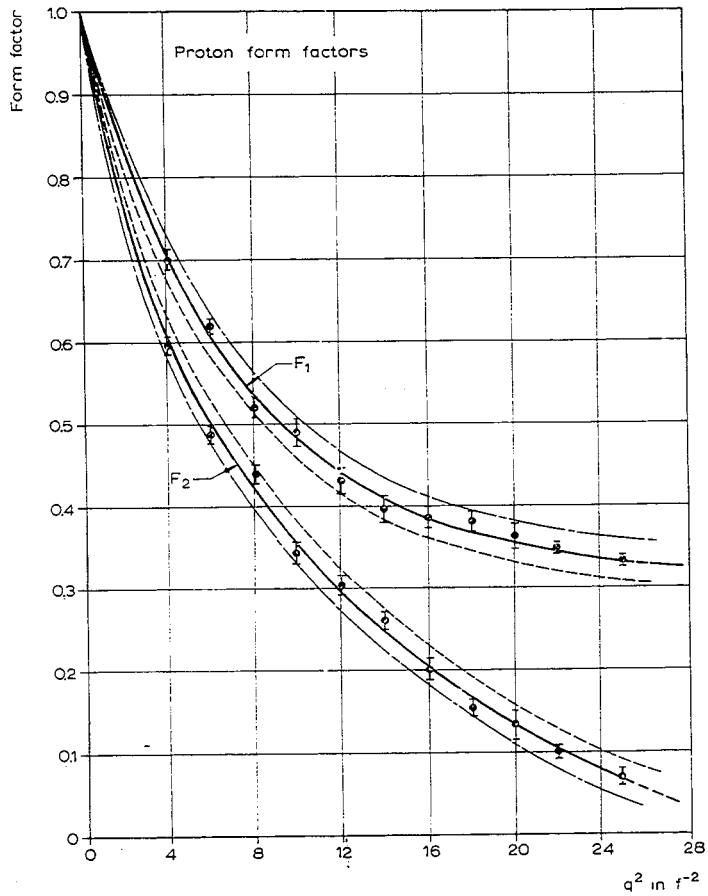


Fig. 10. The most recent Stanford experimental data on the form factors of the proton[17]. There are two dashed curves lying between the *central-value solid experimental lines*. If the error limits are correlated so that they move in opposite directions, as indicated by the *dashed lines*, the corresponding cross sections will still remain consistent with experiment. A similar statement holds for the two *small-dash long-dash curves* lying outside the F_1,F_2 central-value experimental curves. The correlated error question needs additional study but the dashed inner and outer curves are thought to give reasonable error limits of F_1 and F_2.

The early work on the proton was confirmed by subsequent studies at higher energies (~ 600 MeV)[13,14] but these energies were still low enough so that the assumption $F_1 \cong F_2$ could be employed. It was noted in the latter experiments that F_1 was slightly greater than F_2 at values of $q^2 = 4f^{-2}$, where f = fermi = 10^{-13} cm. The value of one fermi corresponds to (197 MeV)$^{-1}$.

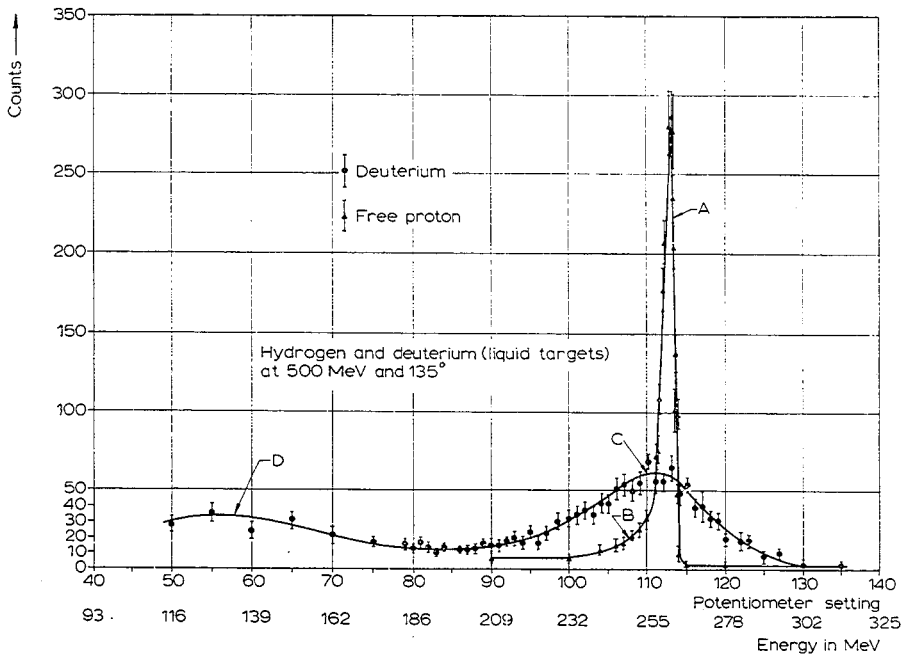


Fig. 11. The experimental comparison of the scattering from the moving proton and neutron in the deuteron (*Curve C*) and the scattering associated with free protons (*Curve A*)[18]. Region *B* shows the bremsstrahlung tail of the proton curve. At *D* are electrons which have been scattered after producing pions in deuterium and also other low-energy electrons. From the scattering data near *C* the form factors of the neutron can be obtained. The proton peak is used for comparison measurements. No correction has been applied in the figure for the different densities of liquid deuterium and liquid hydrogen.

Recently the extension of the experimental measurements to higher energies (~ 1.0 BeV) showed that indeed $F_1 > F_2$[15,16]. The appropriate detailed behavior is shown in Fig. 10, and represents the most recent Stanford experimental data on this subject[17]. The possible theoretical significance of these results will be described below, following brief discussions of first, some tests of the Rosenbluth equation and second, the experimental determinations of the form factors of the neutron.

Various tests of the validity of the Rosenbluth equation were made in these experiments by examining whether F_1 and F_2 are really functions of q alone. In all cases studied for which q^2 was less than $25f^{-2}$ complete consistency in F_1,F_2 values at different energies and angles was observed so that the Rosenbluth equation was checked and found to be valid[17]. At the highest

values reached in these experiments, namely, $q^2 \cong 31\text{f}^{-2}$ the Stanford cross sections could not be combined with the cross sections at the same value of q in recently reported Cornell experiments[24] to give *real* values of F_1 and F_2. If this observation can be confirmed, the possibility exists that quantum electrodynamics may fail at high momentum transfers or that two-proton exchange processes, heretofore neglected, are needed to correct the Rosenbluth equation; or, that some other fundamental aspect of the scattering theory needs improvement. This is an interesting question for the future to decide.

Let us now turn to the question of the neutron. According to relativistic quantum electrodynamics the neutron possesses Dirac and Pauli form factors. Proton and neutron form factors may be referred to respectively as F_{1p}, F_{2p}, F_{1n}, F_{2n}. Static values of the neutron form factors are known to be $F_{1n}(0) = 0$, $F_{2n}(0) = 1.00$. F_{1n} is also known from early neutron-scattering experiments to vary as q^4 at small values of q in an expansion of F_{1n} as a function of q^2. This is commonly referred to by saying that within experimental error, the rms radius of the neutron is zero. Thus the neutron is not only a neutral body from the point of view of electric charge, but has a power expansion of F_{1n} that starts off as a function of q^2 with zero slope! Consequently, it is most difficult to determine F_{1n} (and also F_{2n}) of the neutron. The difficulty is compounded by the experimental fact that neutron targets are obtained only by using the deuteron as a neutron carrier, for free neutrons in large numbers are unobtainable in confined spaces. A neutron is in vigorous motion in the deuteron and this additional complication must be taken into account somehow. It is necessary at this point to introduce a relativistic theory of the deuteron to allow properly for the effects of the motion of the bound neutron. Of course, at the present stage of development of relativity theory, the deuteron problem can be solved only in an approximate way. Hence we can see that there are formidable difficulties which face the experimental elucidation and determination of the form factors of the neutron.

Many of these difficulties were overcome in the work of Yearian and the author who used a difference method to compare the scattering from the deuteron and the proton[18]. These investigators first showed that the neutron could not be represented as a point nucleon and that its magnetic moment was distributed in a manner similar to that of the proton. In Fig. 11, we show the type of data from which such conclusions were drawn. The spread-out deuteron peak shows the effect of the motion of the proton and neutron in

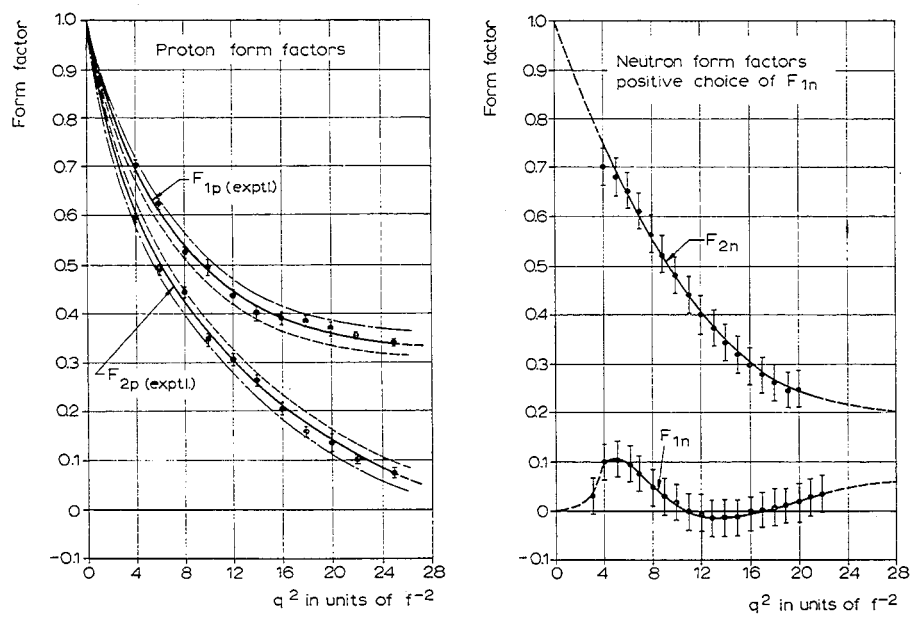


Fig. 12. This figure shows the most recent Stanford results[17,25] for both the neutron and the proton for the positive choice of sign of F_{1n}. The regularity of the neutron curves arises from the fact that the experimental deuteron curves were smoothed before putting the corresponding data into the theoretical formulae from which the form factors are deduced. The four curves of this figure can be fit approximately with dispersion theory or Clementel-Villi curves corresponding to the newly discovered heavy mesons. It is interesting that the newer neutron data agree very well with older results[18] and many of the present conclusions could have been drawn in 1958.

the deuteron and this wide peak may be compared with the sharp peak of the free proton. In the work in which the finite size of the neutron was discovered, the neutron form factor, F_{1n}, was assumed to be approximately zero and F_{2n} had the behavior described above.

It may be noted parenthetically, that it was on the basis of the above results that Nambu[19] postulated the existence of a new heavy neutral meson, now known as ω-meson. Events of the past year have brilliantly confirmed the existence of this meson[20]. A pion-pion resonance (ϱ-meson) responsible for the magnetic behavior of the nucleon form factors was also postulated by Frazer and Fulco[21] on the basis of the above experiments. This resonance was also found recently[22].

The above conclusions about the behavior of the neutron and also the assumption that $F_{1n} \cong 0$, have been confirmed recently[23,24]. More detailed

studies[25], as yet unpublished, support the above description of the neutron form factors. These results are shown in Fig. 12. In Refs. 23 and 24, F_{1_n} was found to be small and positive. However Durand[26] has recently shown that the theory of the deuteron used in the early work to derive the values of the neutron form factors can be improved. When the improved formula is employed the slightly positive values of the form factor, F_{1_n} ,are relatively unaffected in the low q^2 region but in the range $6f^{-2} < q^2 < 20f^{-2}$ the values of F_{1_n} are reduced to approximately zero, within experimental error as in Fig. 12[25]. Because the neutron measurements are so fraught with both experimental and theoretical difficulties we must still regard these new, more accurate results, particularly for F_{1_n}, as preliminary.

Fig. 12 shows the most recent Stanford results for both the proton[17] and the neutron[25]. An ambiguity exists in the choice of sign of F_{1_n}. Fig. 12 shows one choice of the F_{1_n} values and the corresponding F_{2_n} values. Fig. 13 shows the neutron data for the other (negative) choice of F_{1_n} and the corresponding values of F_{2_n} for the negative choice of F_{1_n}. Theoreticians prefer the first choice, but as a purely experimental problem the negative F_{1_n} values must be considered possible until proved untenable. The dashed parts of the curve refer to probable behavior at low q^2 and in the negative F_{1_n} case the steep fall of F_{2_n} would be very surprising and is not expected.

If the first choice of values of F_{1_n} is made, which seems much more likely, an understanding of all the proton and neutron data can be obtained along the lines of the heavy-meson or pion-resonance theory of Bergia, *et al*[27]. An interpretation of the early data in terms of Clementel-Villi form factors, using Yukawa clouds of different ranges and delta functions, was also given by the present author and Herman[23]. These initial and approximate theoretical interpretations are probably correct in principle but incomplete in detail and it now seems likely that it is necessary to add to them the effects on the form factors of a third heavy meson (η-meson)[28]. Such a particle has recently been discovered by A. Pevsner, *et al.*[29]. Its existence was also predicted by Sakurai[30].

Attempts are now being made to fit the data of Fig. 12 in terms of the heavy-meson theory in a way similar to that given in Refs. 23 and 27 but now employing three mesons (ϱ, ω, η) instead of only two. I hesitate to show the results of the studies since the exact mass values of the heavy mesons are not yet definite and small variations of these values affect the relative importance of the mesons in the form factor equations in a sensitive way. Furthermore it would not be surprising to find that new heavy mesons are

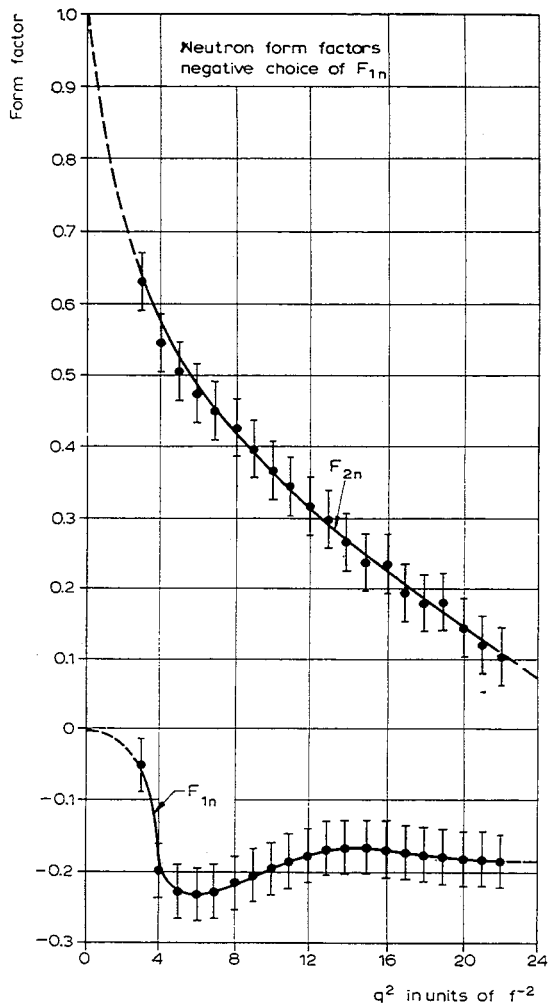


Fig. 13. This figure is similar to the right-hand side of Fig. 12 and gives the set of values of F_{1n} and F_{2n} for the negative choice of F_{1n}. It appears easier at present to fit Fig. 12 with Clementel-Villi curves than Fig. 13.

discovered in the near future, and these might also contribute to the form factors. Suffice it to say that approximate agreement with the data of Fig. 12 can be obtained with the set of three mesons (ϱ, ω, η).

If we now attempt to summarize the recent progress in nucleon structure determinations and in their interpretation, we may say that the proton and neutron are two different aspects of a single entity – the nucleon. The third

component of isotopic spin distinguishes between the two particles. Isotopic form factors can be developed in a well-known way[23] from the proton and neutron form factors. A phenomenological and qualitative interpretation of the nucleon form factors then shows that the same charged mesonic clouds appear in both the neutron and proton. In the proton the clouds add together and in the neutron the clouds cancel, more or less as given in Ref. 23.

It is a bit too early to give the final and definitive details of the mesonic clouds or of their heavy-meson compositions since, as indicated above, such details are now being worked out. However, it is possible, and even likely, that the next year or so should witness a crystallization of the « final » values of the nucleon structure parameters in terms of the models afforded by the new heavy-meson picture of the proton and neutron. The fact that new research is needed in order to clarify this picture is symptomatic of the general problem of the structure of elementary particles.

In concluding this discussion it may be appropriate to return to the theme introduced earlier in the paper and raise the question once again of the deeper, and possibly philosophical, meaning of the term « elementary » particle. As we have seen, the proton and neutron, which were once thought to be elementary particles are now seen to be highly complex bodies. It is almost certain that physicists will subsequently investigate the constituent parts of the proton and neutron – the mesons of one sort or another. What will happen from that point on? One can only guess at future problems and future progress, but my personal conviction is that the search for ever-smaller and ever-more-fundamental particles will go on as long as Man retains the curiosity he has always demonstrated.

The ideas and results presented in this paper represent the work of many individuals. Many of their names are given in the bibliography and I am indebted to them for their important contributions to the subject. I wish to acknowledge my special debt of gratitude to the following individuals who have given me not only invaluable assistance with the many theoretical concepts involved in the fascinating subjects of nuclear and nucleon structure but who have also given me support and encouragement over the last decade: L. I. Schiff, F. Bloch, D. G. Ravenhall, and D. R. Yennie.

1. N. F. Mott, *Proc. Roy. Soc. London*, A 124 (1929) 425; A 135 (1932) 429. (Mott's formula was developed originally in the center of mass frame.)
2. E. Guth, *Anz. Akad. Wiss. Wien, Math.-Naturw. Kl.*, 24 (1934) 299; M. E. Rose, *Phys. Rev.*, 73 (1948) 279; E. Amaldi, G. Fidecaro, and J. Mariani, *Nuovo Cimento*, 7 (1950) 757; J. H. Smith, *Ph. D. Thesis*, Cornell University, 1951 (not published); L. I. Shiff, *Phys. Rev.*, 92 (1953) 988.
3. L. R. B. Elton, *Proc. Phys. Soc. London*, A 63 (1950) 1115; 65 (1952) 481; *Phys. Rev.*, 79 (1950) 412; H. Feshbach, *Phys. Rev.*, 84 (1951) 1206; L. K. Acheson, *Phys. Rev.*, 82 (1951) 488; G. Parzen, *Phys. Rev.*, 80 (1950) 261; 80 (1950) 355.
4. E. M. Lyman, A. O. Hanson, and M. B. Scott, *Phys. Rev.*, 84 (1951) 626.
5. R. Hofstadter, H. R. Fechter, and J. A. McIntyre, *Phys. Rev.*, 91 (1953) 422; *Phys. Rev.*, 92 (1953) 978; R. W. Pidd, C. L. Hammer, and E. C. Raka, *Phys. Rev.*, 92 (1953) 436.
6. D. R. Yennie, D. G. Ravenhall, and R. N. Wilson, *Phys. Rev.*, 92 (1953) 1325; 95 (1954) 500; S. Brenner, G. E. Brown, and L. R. B. Elton, *Phil. Mag.*, 45 (1954) 524; G. E. Brown and L. R. B. Elton, *Phil. Mag.*, 46 (1955) 164; E. Baranger, *Phys. Rev.*, 93 (1954) 1127.
7. V. L. Fitch and J. Rainwater, *Phys. Rev.*, 92 (1953) 789.
8. J. H. Fregeau and R. Hofstadter, *Phys. Rev.*, 99 (1955) 1503.
9. B. Hahn, D. G. Ravenhall, and R. Hofstadter, *Phys. Rev.*, 101 (1956) 1131.
10. R. Hofstadter and R. W. McAllister, *Phys. Rev.*, 98 (1955) 217; R. W. McAllister and R. Hofstadter, *Phys. Rev.*, 102 (1956) 851.
11. M. N. Rosenbluth, *Phys. Rev.*, 79 (1950) 615. See also the reference to a report by L. I. Schiff in this paper, which bears on the problem of proton structure.
12. R. Herman and R. Hofstadter, *High-Energy Electron Scattering Tables*, Stanford University Press, Stanford, California, 1960.
13. E. E. Chambers and R. Hofstadter, *Phys. Rev.*, 103 (1956) 1454.
14. F. Bumiller and R. Hofstadter. See Ref. 12, p. 28; R. Hofstadter, F. Bumiller, and M. R. Yearian, *Revs. Modern Phys.*, 30 (1958) 482.
15. F. Bumiller, M. Croissiaux, and R. Hofstadter, *Phys. Rev. Letters*, 5 (1960) 261; R. Hofstadter, F. Bumiller, and M. Croissiaux, *Phys. Rev. Letters*, 5 (1960) 263; R. Hofstadter, F. Bumiller, and M. Croissiaux, *Proc. 1960 Ann. Intern. Conf. High Energy Physics Rochester*, Interscience Publishers, New York, 1960, pp. 762–766.
16. K. Berkelman, J. M. Cassels, D. N. Olson, and R. R. Wilson, *Proc. 1960 Ann. Intern. Conf. High Energy Physics Rochester*, Interscience Publishers, Inc., New York, 1960, p. 757 ff. Also *Nature*, 188 (1960) 94.
17. F. Bumiller, M. Croissiaux, E. Dally, and R. Hofstadter, *Phys. Rev.*, 124 (1961) 1623.
18. M. R. Yearian and R. Hofstadter, *Phys. Rev.*, 110 (1958) 552; *Phys. Rev.*, 111 (1958) 934; See also, S. Sobottka, *Phys. Rev.*, 118 (1960) 831.
19. Y. Nambu, *Phys. Rev.*, 106 (1957) 1366.
20. B. C. Maglič, L. W. Alvarez, A. H. Rosenfeld, and M. L. Stevenson, *Phys. Rev. Letters*, 7 (1961) 178.
21. W. R. Frazer and J. R. Fulco, *Phys. Rev. Letters*, 2 (1959) 365; *Phys. Rev.*, 117 (1960) 1609. See also, S. Drell, *Proc. 1958 Intern. Conf. High Energy Physics CERN*, pp. 27–33.

22. J. A. Anderson, V. X. Bang, P. G. Burke, D. D. Carmony, and N. Schmitz, *Phys. Rev. Letters*, 6 (1961) 365; A. R. Erwin, R. March, W. D. Walker, and E. West, *ibid*, 628; D. Stonehill, C. Baltay, H. Courant, W. Fickinger, E. C. Fowler, H. Kraybill, J. Sandweiss, J. Sanford, and H. Taft, *ibid*, 624.
23. R. Hofstadter, C. de Vries, and R. Herman, *Phys. Rev. Letters*, 6 (1961) 290; R. Hofstadter and R. Herman, *Phys. Rev. Letters*, 6 (1961) 293.
24. R. M. Littauer, H. F. Schopper, and R. R. Wilson, *Phys. Rev. Letters*, 7 (1961) 141.
25. C. de Vries, R. Hofstadter, and R. Herman, (to be published).
26. L. Durand, III, *Phys. Rev. Letters*, 6 (1961) 631; *Phys. Rev.*, 123 (1961) 1393. See also A. Goldberg, *Phys. Rev.*, 112 (1958) 618.
27. S. Bergia, A. Stanghellini, S. Fubini, and C. Villi, *Phys. Rev. Letters*, 6 (1961) 367; S. Bergia and A. Stanghellini, *Nuovo Cimento*, 21 (1961) 155.
28. C. de Vries, R. Hofstadter, R. Herman, and S. Krasner, *Proc. Aix-en-Provence Conf. Elementary Particles*, Sept. 1961.
29. 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* (to be published). (I wish to thank Professor Pevsner for communicating his results to me before publication.)
30. J. J. Sakurai, *Phys. Rev. Letters*, 7 (1961) 355 and private communication.

Biography

Robert Hofstadter, Professor of Physics at Stanford University, was born in New York, N.Y., of parents Louis Hofstadter and Henrietta Koenigsberg, on February 5, 1915.

Hofstadter attended elementary and high schools in New York City, and was graduated in 1935 from the College of the City of New York with the B.S. degree, *magna cum laude*.

On graduation from college Hofstadter received the Kenyon Prize in Mathematics and Physics, and a little later the Coffin Fellowship, awarded by the General Electric Company. He went to graduate school at Princeton University where he studied physics from 1935–1938, and received both the M.A. and Ph.D. degrees in 1938 from that institution. His Ph.D. work was concerned with infrared spectra of simple organic molecules, and in particular, with the partial elucidation of the structure of the now well-known « hydrogen bond ». In 1938–1939 he was awarded a Procter Fellowship at Princeton University for postdoctoral work, at which time he began a study of photoconductivity in willemite crystals. This work led to the discovery, with R. Herman, of the warm-up dark currents which demonstrated the existence of trapping states in crystals. In 1939 Hofstadter received the Harrison Fellowship at the University of Pennsylvania where he helped to construct a large Van de Graaff machine for nuclear research. At Pennsylvania he first met L. I. Schiff, who has been a friend and colleague for many years.

During the war years Hofstadter worked first at the National Bureau of Standards and later at the Norden Laboratory Corporation. He left industry at the end of the war to become Assistant Professor of Physics at Princeton University. At Princeton he carried out research on crystal conduction counters, on the Compton effect, and on scintillation counters. In 1948 he discovered that sodium iodide, activated by thallium, made an excellent scintillation counter. In 1950, with J. A. McIntyre, he found that well-formed crystals of this material provided remarkable energy-measuring devices for gamma rays and energetic particles and thus could be used as spectrometers in addition to gamma-ray and particle counters of high efficiency.

In 1950 Hofstadter left Princeton to become Associate Professor of Physics at Stanford University where he initiated a program on the scattering of energetic electrons from the linear accelerator, invented by W. W. Hansen, which was then under construction. While building equipment for the electron-scattering experiments, he continued working on scintillation counters and developed new detectors for neutrons and X-rays. High-speed inorganic (CsF) and useful Čerenkov (TlCl) counters were discovered at Stanford. Other studies carried out in the early years at Stanford were concerned with cosmic rays and with cascade showers generated by high-speed electrons.

After 1953 electron-scattering measurements became Hofstadter's principal interest. With students and colleagues he investigated the charge distribution in atomic nuclei and afterwards the charge and magnetic moment distributions in the proton and neutron. The electron-scattering method was used to find the size and surface thickness parameters of nuclei. Many of the principal results on the proton and neutron were obtained in the years 1954–1957. Since 1957 emphasis in the research program has been placed on making more precise studies of the nucleon form factors. This work is still in progress.

Hofstadter was elected to the National Academy of Sciences (U.S.A.) in 1958 and was named California Scientist of the Year in 1959. He has also been a Guggenheim Fellow (1958–1959) and spent one year at CERN in Geneva, Switzerland, on sabbatical leave.

In 1942 he married Nancy Givan of Baltimore, Maryland, and they have a son, Douglas, and two daughters, Laura and Mary.

RUDOLF L. MÖSSBAUER

Recoilless nuclear resonance absorption of gamma radiation

Nobel Lecture, December 11, 1961

It is a high distinction to be permitted to address you on the subject of recoilless nuclear resonance absorption of gamma radiation. The methods used in this special branch of experimental physics have recently found acceptance in many areas of science. I take the liberty to confine myself essentially to the work which I was able to carry out in the years 1955 to 1958 at the Max Planck Institute in Heidelberg, and which finally led to establishment of the field of recoilless nuclear resonance absorption. Many investigators shared in the preparations of the basis for the research we are concerned with in this lecture. As early as the middle of the last century Stokes observed, in the case of fluorite, the phenomenon now known as fluorescence – namely, that solids, liquids, and gases under certain conditions partially absorb incident electromagnetic radiation which immediately is reradiated. A special case is the so-called resonance fluorescence, a phenomenon in which the re-emitted and the incident radiation both are of the same wavelength. The resonance fluorescence of the yellow D lines of sodium in sodium vapour is a particularly notable and exhaustively studied example. In this optical type of resonance fluorescence, light sources are used in which the atoms undergo transitions from excited states to their ground states (Fig. 1). The light quanta emitted in these transitions ($A \rightarrow B$) are used to initiate the inverse process of resonance absorption in the atoms of an absorber which are identical with the radiating atoms. The atoms of the absorber undergo a transition here from the ground state (B) to the excited state (A), from which they again return to the ground state, after a certain time delay, by emission of fluorescent light.

As early as 1929, Kuhn[1] had expressed the opinion that the resonance absorption of gamma rays should constitute the nuclear physics analogue to this optical resonance fluorescence. Here, a radioactive source should replace the optical light source. The gamma rays emitted by this source should be able to initiate the inverse process of nuclear resonance absorption in an absorber composed of nuclei of the same type as those decaying in the source.

Again, the scheme of Fig. 1 would hold here, but the radiative transitions would now take place between nuclear states. Nevertheless, all attempts in the next two decades to find this nuclear resonance absorption proved fruitless. Before I can approach the subject of my talk, it is appropriate to consider the reasons why the discovery of nuclear resonance absorption was so long delayed.

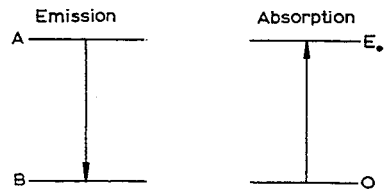


Fig. 1. Scheme of resonance absorption.

For simplicity, we shall first consider a nuclear transition of a free nucleus at rest. The gamma quantum emitted in the transition $A \rightarrow B$ imparts a recoil momentum $\vec{p}$ to the emitting nucleus and consequently a kinetic energy ΔE, which is given by

$$\Delta E = \vec{p}^2/2M = E_0^2/2Mc^2 \qquad (1)$$

where M is the mass of the nucleus and c is the velocity of light. The energy liberated in this nuclear transition is divided, in accordance with the law of conservation of momentum, the larger part being carried away by the emitted quantum, the other part going to the emitting nucleus in the form of recoil energy. This recoil-energy loss of the quantum has the consequence that the emission line does not appear at the position of the transition energy E_0 but is displaced to lower energy by an amount ΔE (Fig. 2). The absorption line, on the contrary, is displaced to a higher energy by the same amount ΔE, because in order for the process of resonance absorption to occur, a quantum must provide, in addition to the transition energy E_0, the amount of energy ΔE which is taken up by the absorbing nucleus in the form of a recoil kinetic energy. Typical values for the line shifts ΔE lie in the range from 10^{-2} eV to 10^2 eV; they are therefore very small in comparison with the energies of the gamma quanta, which frequently are of the order of magnitude of millions of electron volts.

Since there is an uncertainty in the energies of the individual excited states

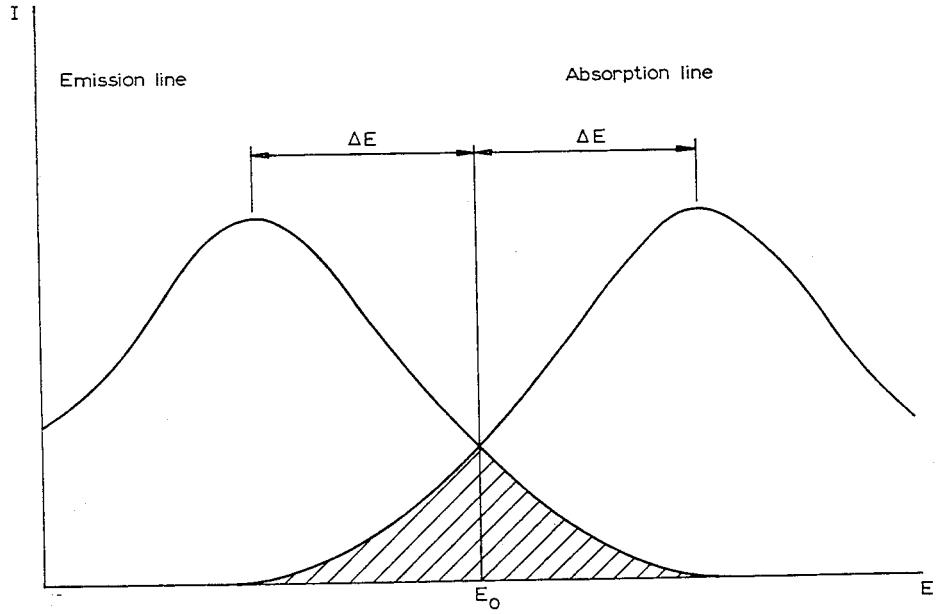


Fig. 2. Position and shape of the emission and absorption lines of a free nucleus (shown in the case of the 129-keV transition in ^{191}Ir for $T = 300°K$).

of the nuclei, the lines associated with transitions between an excited state and the ground state have a certain minimum width. This so-called natural width Γ is, according to the Heisenberg uncertainty principle, connected with the lifetime τ of an excited nuclear state by the relation $\Gamma\tau = \hbar$. The usual values for the lifetimes τ of the low-lying excited nuclear states lie in the range from 10^{-7} to 10^{-11} second, corresponding to an interval of 10^{-8} to 10^{-4} eV for the natural line widths appearing in ground-state transitions. Such extraordinarily sharp lines exhibiting the natural line width are normally not observed. Rather, a series of side effects always exist, which lead to considerable broadening such that the line widths ordinarily associated with low-energy gamma transitions exceed the natural minimum width by many orders of magnitude. The most important broadening mechanism is the thermal motion of the nuclei in the source and in the absorber; this leads to Doppler shifts in the energy of the gamma quanta and therefore produces corresponding line broadenings. Even a temperature drop to absolute zero does not extinguish this thermal broadening, since the existence of zero-point energy at absolute zero frequently produces, at least in solids, line

widths of the same order of magnitude as the ones existing at room temperature.

Usually the line shifts ΔE are large in comparison to the thermal line widths, and they are also always very large in comparison to the natural line widths associated with the low-energy nuclear transitions we are concerned with here. As a consequence, the energy of an emitted quantum is usually too small for the inverse process of resonance absorption to be carried out, or in other words, the probability of occurrence of nuclear resonance absorption is so small that the process escapes detection. Therefore the long and unsuccessful search for nuclear resonance absorption is to be blamed on the high recoil-energy losses of the gamma quanta.

On the other hand, entirely different conditions hold for optical resonance absorption. There, because of the much lower energies of the light quanta, the recoil-energy losses produced by light quanta are small in comparison with the line widths. Emission and absorption lines therefore overlap in an ideal manner, the resonance condition is satisfied, and the optical effect is, at least in principle, easily observable.

The unsatisfactory situation with respect to nuclear resonance absorption first changed in 1951, when Moon[2] succeeded in demonstrating the effect for the first time, by an ingenious experiment. The fundamental idea of his experiment was that of compensating for the recoil-energy losses of the gamma quanta: the radioactive source used in the experiment was moved at a suitably high velocity toward the absorber or scatterer. The displacement of the emission line toward higher energies achieved in this way through the Doppler effect produced a measurable nuclear fluorescence effect.

After the existence of nuclear resonance fluorescence had been experimentally proved, a number of methods were developed which made it possible to observe nuclear resonance absorption in various nuclei. In all these methods for achieving measurable nuclear resonance effects the recoil-energy loss associated with gamma emission or absorption was compensated for in one way or another by the Doppler effect.

At this point, let me speak of my esteemed teacher, Heinz Maier-Leibnitz, who in 1953 directed my attention to this newly advancing field of nuclear resonance fluorescence, and who stimulated me to turn to this area of research. He was, also, the one who made it possible for me to conduct my research throughout the years 1955 to 1958 at the Heidelberg Max Planck Institute, in an undisturbed and fruitful atmosphere – research which finally led to the discovery of recoilless nuclear resonance absorption. I want to express

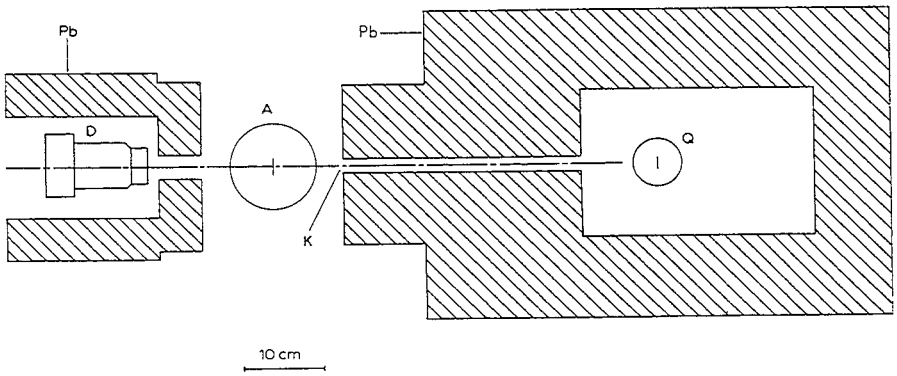


Fig. 3. Experimental arrangement. *A*, absorber-cryostat; *Q*, cryostat with source; *D*, scintillation detector; *K*, lead collimator.

my warmest thanks to my esteemed teacher for his efforts on my behalf.

The method which I shall now proceed to discuss differs fundamentally from the methods described above in that it attacks the problem of recoil-energy loss at its root in a manner which, in general, insures the complete elimination of this energy loss. The basic feature of this method is that the nuclei in the source and absorber are bound in crystals. The experiments described in the following paragraphs exclusively employed radioactive sources which emitted the 129-keV gamma line leading to the ground state in ^{191}Ir.

The first experiment[3] aimed at measuring the lifetime of the 129-keV state in ^{191}Ir by utilizing methods of nuclear resonance absorption known at that time. The experimental set-up used for this purpose is shown schematically in Fig. 3. A method first employed by Malmfors[4] appeared to be especially suitable for the planned measurement. In this method, a broadening of the emission or absorption line, leading to a corresponding increase in the degree of overlap of the two lines, is achieved by increasing the temperature. If the relative shift of the emission and the absorption lines resulting from the recoil-energy losses is only of the order of magnitude of the line widths, a temperature increase leads, under favourable conditions, to a measurable nuclear absorption effect. In the case of the 129-keV transition in ^{191}Ir, there is considerable overlap of the two lines even at room temperature as a consequence of the small energies of the quanta and the small line shifts resulting therefrom (see Fig. 2). In this case, not only an increase but also a decrease in temperature can result in a measurable change in the nuclear absorption.

My decision between these two possibilities was made in favour of a temperature decrease. It was motivated essentially by the consideration that at low temperature, effects of chemical binding would be more likely than at elevated temperatures. This hypothesis was vindicated in an unexpected way during the course of the experiments. The simultaneous cooling of the source and the absorber with liquid air led to inexplicable results, for which I first blamed effects associated in some way with the cooling of the absorber. In order to eliminate these unwanted side effects, I finally left the absorber at room temperature and cooled only the source. In very tedious experiments, which demanded extremely stable apparatus, a small decrease in the absorption with respect to the value at room temperature was in fact obtained – a result consistent with my expectations. The evaluation of these measurements finally led to the determination of the lifetime sought for.

In a second series of experiments I attempted to explain the side effects which had appeared in the simultaneous cooling of the source and the absorber during the earlier experiments. The result of this attempt was striking:

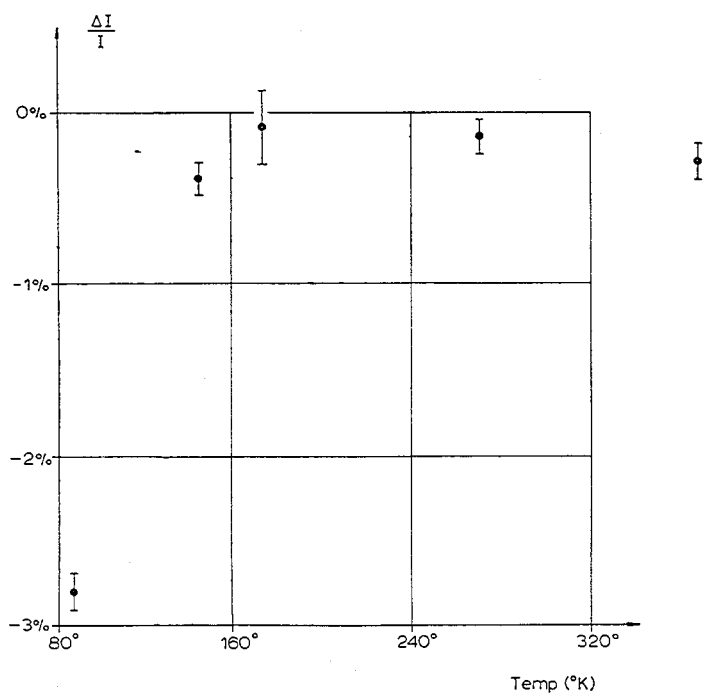


Fig. 4. Temperature dependence of the absorption. Relative intensity change $\Delta I/I$ in comparison with that of a non-resonant absorber.

instead of the decrease expected, a strong increase in the absorption clearly manifested itself when the absorber was cooled. This result was in complete contradiction to the theoretical expectation. The observed temperature dependence of this absorption is shown in Fig. 4.

In considering the possible sources of the anomalous resonance effect, I now began to subject the hypothesis of the existing theory to a critical examination.

The views originally held as to the shape and energy of the emission and absorption lines were based on the assumption that the emitting and absorbing nuclei can be treated as free particles. It was therefore natural to modify this assumption, taking into account the fact that source and absorber were each used in crystalline form. Therefore, I first attempted to explain the observed anomalous resonance absorption by assuming that the recoil momentum was not transferred to the single nucleus. It should rather be transferred to an assembly of nuclei or atoms which include nearest or next-nearest neighbours surrounding the nucleus under consideration. After the failure of this and other attempted explanations, based on a purely classical point of view, I turned my attention to a quantum-mechanical treatment of the problem.

Let me here introduce some concepts from crystal physics, by means of which I will develop the reasoning that finally led to a solution of the problem. All the internal motions of the particles forming the crystal lattice can be described in terms of a superposition of a large number of characteristic vibrations whose distribution is called the frequency spectrum of the crystal. The nature of the binding of the particles forming the crystal determines the structure of the lattice vibrational spectrum, this spectrum often being very complicated. Nevertheless, the substitution, for the true vibrational spectrum of the lattice, of a much simpler frequency distribution often suffices for qualitative considerations.

This first, still purely classical picture of the internal motions in a crystal corresponds, in the quantum-mechanical description, to a system of uncoupled harmonic oscillators with quantized energy states. While the vibrational spectrum describes only the distribution of the fundamental frequencies of these oscillators, the temperature determines the so-called occupation numbers of these crystal oscillators; these occupation numbers simply tell which of the possible energy states the particular oscillators occupy.

The recoil energy appearing in the emission or absorption of a quantum by a nucleus bound in a crystal is taken up by the crystal partly in the form

of translational energy and partly in the form of internal energy. The resultant increase in translational energy is always negligible because of the enormous mass of the crystal as a whole in comparison to the mass of a single nucleus. An increase in the internal energy leads to changes in the occupation numbers of the individual crystal oscillators. Because of the quantization of the oscillator energies, the crystal can absorb the recoil energy only in discrete amounts. The nuclear transitions of the bound nuclei are normally accompanied by simultaneous transitions of the crystal oscillators. Thus, for example, a gamma quantum can be emitted and simultaneously one of the crystal oscillators can undergo a transition to a neighbouring energy state. Likewise, a gamma emission process can be accompanied by simultaneous transitions of two crystal oscillators, and so forth. As a consequence of the quantization of the oscillators, there also exists in principle the possibility that the gamma transition takes place with none of the crystal oscillators changing their states.

The problem now was to compute the probabilities of the various processes. Significant was the calculation of the probability of nuclear transitions leaving the lattice state unchanged – that is, transitions in which no recoil energy is transferred to the lattice in the form of internal energy. Similar problems had already been solved much earlier. The coherent scattering of X-rays from crystals had been known for decades; in that case a momentum transfer to the reflecting lattice takes place without simultaneous transfer of internal energy to the crystal. And for a long time the analogous problem of the elastic scattering of slow neutrons from crystals had been thoroughly studied, both experimentally and theoretically. Lamb[5] had, as early as 1939, developed a theory for the resonance capture of slow neutrons in crystals. This theoretical work was somewhat premature, inasmuch as Lamb assumed considerably smaller values for the widths of the neutron lines than were observed in later experiments. For this reason, this extraordinarily beautiful work was of no practical significance in the area of application originally intended. It remained only to apply this Lamb's theory to the analogous problem of the resonance absorption of gamma radiation. This indeed enabled me to show that under the conditions chosen in the experiments described above there exists a high probability of nuclear transitions with no simultaneous change of the lattice state. Since these nuclear transitions are not associated with any energy losses caused by recoil phenomena, I shall in the following discussion characterize these particular transitions as « recoilless » transitions, and the lines associated with such transitions, as « recoilless

lines». In such a recoilless emission process, the entire excitation energy is therefore transferred to the emitted quantum, and the corresponding situation holds for the recoilless absorption. Here the notation « recoilless » relates only to recoil energy transferred in a nuclear transition, and not to the transferred momentum. The value of this transferred momentum is determined by the energy of the gamma quantum and is essentially a constant, independent of any change in the internal state of motion of the crystal. This momentum is, therefore, transferred to the lattice in all emission or absorption processes, even in the recoilless processes. It is always taken up by the crystal as a whole, and therefore the corresponding translational velocity is negligibly small.

What are the conditions under which the recoilless nuclear resonance absorption can be observed? In answering this question, I wish to develop here, without presenting the mathematical formulation of the theory, a detailed picture of the processes which take place in radiative transitions in nuclei bound in crystals. The recoil-energy loss that occurs in a nuclear transition of a free nucleus is given by Eq. (1). It can be shown that in a transition of a nucleus bound in a crystal, Eq. (1) is no longer valid for the individual process but holds in the means over many processes; that is, instead of Eq. (1), we now have

$$\overline{\Delta E} = E_0^2/2Mc^2 \qquad (2)$$

Let us now consider a very simple model, in which we describe the vibrational state of the crystal by a single frequency ω, the so-called Einstein frequency. It is instructive to consider the two limiting cases, in which the mean recoil energy is either large or small in comparison to the transition energy of the Einstein oscillator:

$$\overline{\Delta E} > \hbar\omega \quad \text{(case 1)}$$
$$\overline{\Delta E} < \hbar\omega \quad \text{(case 2)}$$

In case 1, many oscillator transitions are required to take up the energy contribution $\overline{\Delta E}$ in the lattice. The nuclear processes will therefore in general be accompanied by simultaneous transitions of many oscillators. The probability of a nuclear transition taking place without any oscillator transition – that is, the probability of a recoilless process – is correspondingly small. The situation is entirely different in case 2. Here it is immediately evident that

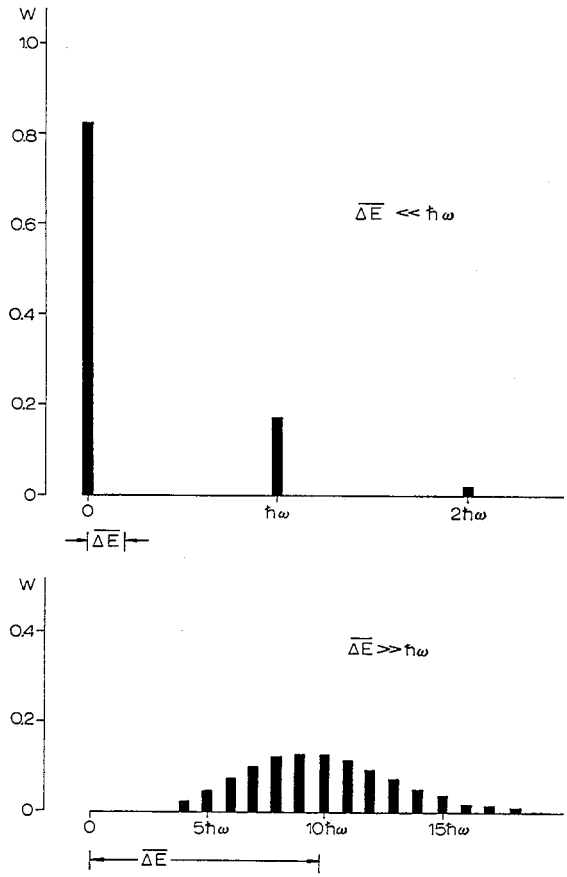


Fig. 5. Relative probability for a gamma transition associated with simultaneous transitions of 0, 1, 2, . . . n crystal oscillators, for two different values of $\overline{\Delta E}$ for the temperature $T = 0°K$.

a nuclear transition which is accompanied by an oscillator transition occurs relatively seldom, leading under these circumstances to a high probability for recoilless processes. The probabilities, under these conditions, for the occurrence of gamma emission processes accompanied by 0, 1, 2, . . . n oscillator transitions are shown in Fig. 5 for these two cases.

For qualitative considerations this simple picture can well be applied to the case of the real crystal. The frequency spectrum of the real crystal exhibits an increasingly high oscillator density at high frequencies. It is sufficient, for this simplified consideration, to replace the Einstein frequency by the upper frequency of the vibrational spectrum of the real crystal. This

limiting frequency ω_g is related, approximately, to the characteristic temperature θ of the crystal by the equation $\hbar\omega_g = k\theta$.

The essential condition for a high probability of recoilless nuclear transitions now has the form

$$E_0^2/2Mc^2 < k\theta \tag{3}$$

The condition given here is quite restrictive because it limits the observation of recoilless resonance absorption to nuclear transitions of relatively low energy; the upper limit lies at about 150 keV.

If $\overline{\Delta E} = E_0^2/2Mc^2$ is small in comparison to the upper energy limit of the frequency spectrum, the percentage of recoilless processes that occur is high even at room temperature. However, if $\overline{\Delta E}$ is about equal to the upper energy limit of the frequency spectrum, then the probability of transitions of the crystal oscillators must be correspondingly reduced by the use of low temperatures in order to arrive at measurable effects.

After the appearance of the observed strong resonance absorption in ^{191}Ir had been traced back to the phenomenon of recoilless nuclear resonance absorption, the next step was to compute the probability of the effect in a general form. This probability, also known as the Debye-Waller factor, in analogy with the terminology used in X-ray scattering, is, as I have already pointed out, strongly dependent on the temperature and the energy of the nuclear transition. This dependence is illustrated in Fig. 6 by two examples.

The shape and location of the emission and absorption lines, as shown in Fig. 2 for the case of a free nucleus, are modified considerably by the influence of the chemical binding. While the centres of the lines as given by Eq. (2) are retained, each of them shows a complicated structure which reflects different single and multiple oscillator transitions, hence the notion « line » is applied here in a more generalized sense.

The most interesting prediction of the theory is the appearance of a strong line with the natural line width appearing in the structure of both « lines » at the position of the transition energy E_o. This line represents the recoilless processes. The strong prominence of these lines with the natural line width within the total line structure is not so surprising when one considers that all the recoilless processes in the emission and absorption spectra appear within an energy range of the order of magnitude of the natural line width. The gamma transitions associated with the oscillator transitions appear, on the other hand, in an energy range of the order of magnitude of the Debye

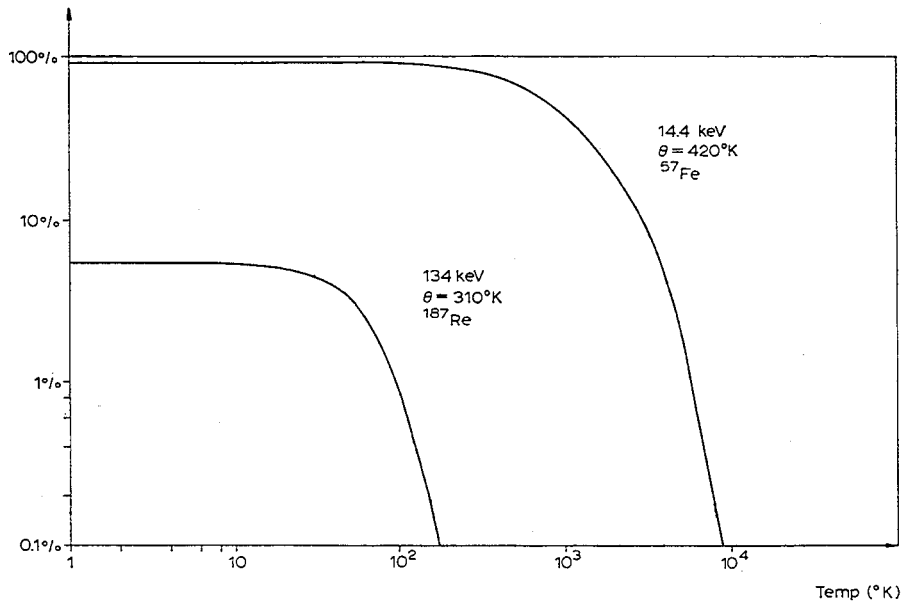


Fig. 6. Fractions of recoil-free nuclear transitions (Debye-Waller factors) in ^{57}Fe and ^{187}Re, shown as functions of the temperature.

temperature of the crystal; this energy range is always broader than the natural line width by many orders of magnitude.

All these considerations provided a very plausible explanation of the origin of the observed resonance absorption in ^{191}Ir at low temperatures. The agreement between the experimentally observed temperature dependence of the absorption (Fig. 4) and the theoretically computed dependence was satisfactory when one considered that in the first calculation the Debye approximation was used for the frequency spectrum because the actual frequency spectra of the crystals used were not known. Notwithstanding this qualitative agreement between experiment and theory, the situation still appeared to be somewhat unsatisfactory. The theory predicted that under the prescribed conditions lines of natural width should appear in the structure of the emission and absorption lines at the value of the transition energy E_o.

While the assumption of the existence of these lines made it possible to explain the experiments thus far carried out, the keystone was evidently still missing – namely, direct experimental proof of the existence of these lines and, especially, a demonstration that their widths were indeed the natural

line widths. It was necessary to find a detector which had the necessary energy resolution to measure the profiles of these extremely sharp lines. The use of conventional detectors was excluded from the start. For example, the scintillation detectors frequently used for gamma-ray measurements in the relevant energy region have resolutions of the order of magnitude of 10^4 eV, while the natural line widths of the observed lines were only about 10^{-6} eV. The possibility of using the atomic nuclei themselves as detectors was suggested as a way out of this situation. As was shown above, an essential prediction of the theory was that the recoilless lines in both the emission and the absorption spectra should appear at the same position – that is, at the value of the transition energy.

Both lines would, therefore, overlap completely, this being the reason for the strong absorption effect. If one could succeed in partially removing the perfect overlap by a relative displacement of the lines, the absorption effect of the recoilless lines should disappear correspondingly.

In the experiment carried out to test the validity of this prediction, the idea was to accomplish the relative shift of the lines by means of the Doppler effect, so that a relative velocity would be imparted to the nuclei of the source with respect to the nuclei of the absorber. Here we had a sort of reversal of the experiment carried out by Moon. Whereas in that experiment the resonance condition destroyed by the recoil-energy losses was regained

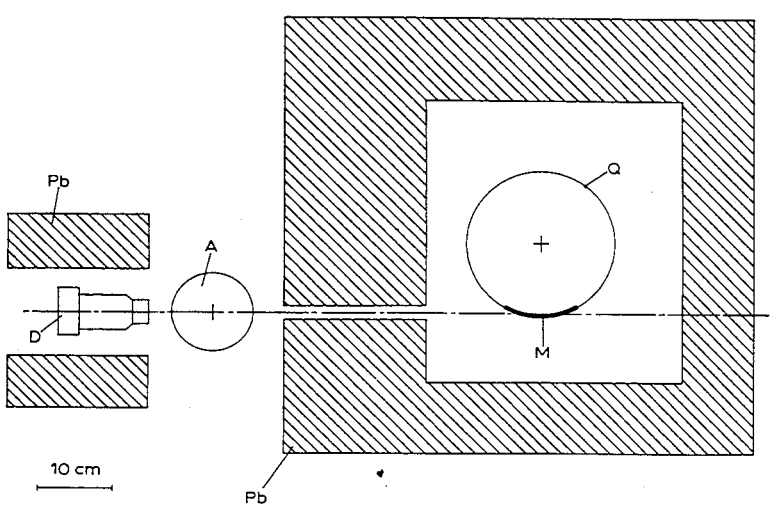


Fig. 7. Experimental arrangement. *A*, absorber-cryostat; *Q*, rotating cryostat with source; *D*, scintillation detector.

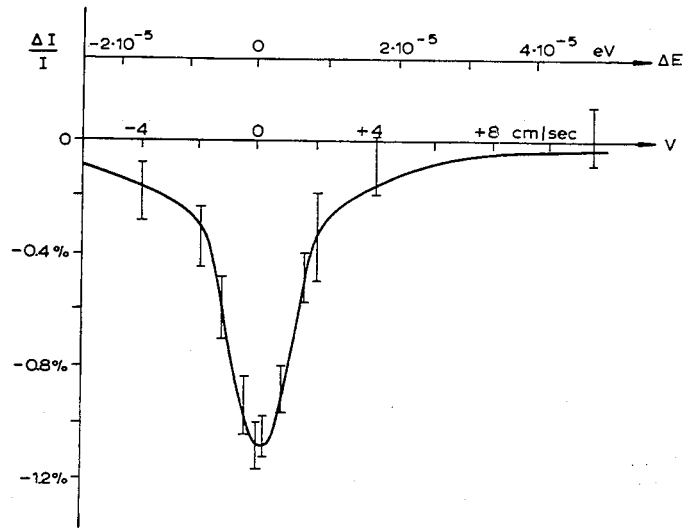


Fig. 8. Relative intensity ratio $\Delta I/I$ of gamma radiation measured behind the resonant iridium absorber, in comparison with intensities measured behind a nonresonant absorber.

by the application of an appropriate relative velocity, here the resonance condition fulfilled in the experiment was to be destroyed through the application of a relative velocity. And yet there was an essential difference between this and Moon's experiment. There, the width of the lines that were displaced relative to one another was determined by the thermal motion of the nuclei in the source and absorber; here, the line widths were sharper by four orders of magnitude. This made it possible to shift them by applying velocities smaller by four orders of magnitude. The indicated velocities were in the region of centimeters per second.

Fig. 7 shows the experimental arrangement[6]. For simplicity, I decided to move the source by means of a turn-table. Only the part of the rotational motion marked by the heavy line in Fig. 7 was used for the measurement – namely, that part in which the source was moving relative to the absorber with approximately constant velocity. The intensity at the detector was measured as a function of the relative velocity between the source and the absorber. Since the preparation of the conical-gear assembly necessary for adjusting the various velocities caused a disagreeable delay in this experiment which was so exciting for me, I took advantage of the existence in Germany of a highly developed industry for the production of mechanical toys. A

day spent in the Heidelberg toy shops contributed materially to the acceleration of the work.

Fig. 8 shows the result of this experiment, a result which was just what had been expected. As the figure demonstrates, a maximum resonance absorption was actually present at zero relative velocity as a result of the complete superposition of the recoilless emission and absorption lines; therefore, minimal radiation intensity passing through the absorber was observed in the detector. With increasing relative velocity the emission line was shifted to higher or lower energies, the resonance absorption decreased, and the observed intensity correspondingly increased. The necessary relative velocities were manifestly only of the order of centimeters per second. Since the experiment consisted essentially of producing a shift of an emission line of width Γ relative to an absorption line of width Γ, the observed line possessed a width which, with a small correction, was equal to $2\,\Gamma$. It was especially satisfying that the line width thus obtained agreed with the width determined in the first experiment[3] under much more difficult conditions. While absorption effects of the order of 1 per cent were observed in the second experiment, an effect of the order of a hundredth of 1 per cent had been achieved in the earlier work. Thus, direct proof of the existence of recoilless absorption was achieved.

The significance of the new method was immediately apparent, although not all of its consequences were immediately realized. With 4.6×10^{-6} eV the line width observed in the 129 keV gamma transition of ^{191}Ir was already smaller than the usual thermal line widths by many orders of magnitude. Let us define the energy resolving power $E/\delta E$ by the ratio of the energy E_0 of the nuclear transition and the natural width Γ of the line; that is, $E/\delta E = E_0/\Gamma$. Then we obtain in the present case $E/\delta E = 2.8 \times 10^{10}$. The actual energy resolution achieved was already much greater in this experiment, since energy shifts amounting to a small fraction of the natural line width were actually observed. Since then, it has become possible, by utilizing other nuclear isotopes, to improve the obtainable energy resolution once more by several orders of magnitude. It is this property of the recoilless nuclear resonance absorption – namely, that it is possible by this means to measure extraordinarily small energy differences between two systems – which gave the method its significance and opened up a broad field of possible applications.

Thus, the extraordinary sharpness of the recoilless gamma lines brought direct investigation of the hyperfine structure of nuclear transitions within

the range of possibility. As a rule, atomic nuclei possess electric and magnetic moments in their various excited states. The interaction of these moments with internal or external fields leads to a splitting of nuclear levels into a number of states that are very close to one another. This hyperfine structure normally remains hidden in gamma lines, since the thermal width of a gamma line is always very large in comparison to the spacing of the hyperfine levels of the nuclear state. When the distances between the individual hyperfine components are larger than the natural widths of the gamma lines, as is frequently the case, and when the conditions necessary for the observation of recoilless absorption are fulfilled, this method makes possible direct measurements of the hyperfine structure of both participating nuclear states. Instead of a single line with the natural line width, a whole set of lines now appears in the emission spectrum, corresponding to recoilless transitions between the hyperfine levels of the nuclear excited and ground states. The same situation holds for the absorption spectrum. Studies of this type yield predictions on the magnetic dipole and electric quadrupole moments of the nuclear states involved, as well as predictions on the magnetic fields and gradients of the electric field prevailing at the site of the nucleus. The special promise of such measurements lies primarily in the possibility of obtaining information on the hyperfine splitting of excited nuclear states. In fact, several moments of excited nuclear states have been determined in this way in various laboratories. However, in our laboratory and in laboratories of many others now working in this region, this method is used mainly for studying the internal fields existing at the site of the nucleus; such studies have already led to a series of interesting results.

In addition to measurement of the fields located in crystals at nuclear sites and to measurement of the moments of excited nuclear states, studies of a number of important effects have been made during the past two years in a large number of laboratories. The observation of these effects was made possible by means of even sharper nuclear transitions, especially that of the 14.4-keV transition in ^{57}Fe.

Particular mention should be made here of the beautiful measurements of the energy shift of radiation quanta in the gravitational field of the earth[7], the observation of the second-order Doppler effect, and the measurements of the isomeric shift. However, discussion of these and of a whole series of other effects which have been studied by means of the method of recoilless nuclear resonance lies outside the framework of this address.

The interpretation of the formalism which underlies the quantitative de-

scription of recoilless nuclear resonance absorption has led to extremely active discussions. The question particularly raised has been whether the effect can be explained classically and, in this connection, whether the momentum transfer takes place as a continuous process during the lifetime τ of the excited nuclear state or as a spontaneous process in the sense of quantum electrodynamics. The question raised here is closely connected with the problems encountered in the dualistic description of radiation as a wave process or as a stream of free particles. Notwithstanding the fact that many details of the recoilless resonance absorption can be described by classical models, I should characterize this effect as a specifically quantum-mechanical one. In particular, it can be shown by the mathematical formulation of the theory that the momentum transfer takes place spontaneously. This can be demonstrated experimentally, not by an individual process but, rather, by the measurement of certain integral quantities, as, for example, the Debye-Waller factor.

In conclusion, let me now say a few words on the limits of the methods described. An upper limit for the usable natural line widths has certainly been given when the widths approach the width of the vibrational spectrum of the crystal oscillators.

In this particular case it becomes impossible to distinguish clearly between nuclear processes which are simultaneously accompanied by oscillator transitions and those which are not. However, this would require nuclear lifetimes of less than 10^{-13} second, which do not occur in the nuclear levels available for this type of experiment. Therefore this upper limit is both uninteresting and unrealistic.

More difficult is the question of a lower limit for the available line widths. To go beyond the limit so far attained of about 10^{-8} eV for the natural line widths is quite possible in principle. There are, however, a number of factors which have delayed extension of the method into the region of higher sensitivity. On the one hand, there are very few nuclei of stable isotopes whose first excited states possess the desired lifetime of greater than 10^{-7} second. On the other hand, all the previously mentioned side effects of such sharp lines play a dominant role; small disturbances of the crystal symmetry and small contaminations quickly lead to individual shifts in the nuclear states, and these shifts, as a group, produce a very considerable broadening of the extremely sharp lines. In this way the resonance condition is so far violated that the lines are not observed. However, there is a well-founded view that still more sharply defined nuclear transitions will be available before long.

These should lead to multiplication of the possibilities for applying the method of recoilless absorption. We may therefore hope that this young branch of physics stands only at its threshold, and that it will be developed in the future, not only to extend the application of existing knowledge but to make possible new advances in the exciting world of unknown phenomena and effects.

1. W. Kuhn, *Phil. Mag.*, 8 (1929) 625.
2. P. B. Moon, *Proc. Phys. Soc. (London)*, 64 (1951) 76.
3. R. L. Mössbauer, *Z. Physik*, 151 (1958) 124.
4. K. G. Malmfors, *Arkiv Fysik*, 6 (1953) 49.
5. W. E. Lamb, Jr., *Phys. Rev.*, 55 (1939) 190.
6. R. L. Mössbauer, *Naturwiss.*, 45 (1958) 538; *Z. Naturforsch.*, 14a (1959) 211.
7. R. V. Pound and G. A. Rebka, Jr., *Phys. Rev. Letters*, 4 (1960) 337. A complete bibliography of all pertinent researches is included in H. Frauenfelder, *The Mössbauer Effect*, to be published in the series *Frontiers in Physics*, Benjamin, New York.

Biography

Rudolf Ludwig Mössbauer was born in Munich on the 31st of January 1929, the son of Ludwig Mössbauer and his wife Erna, *née* Ernst. He was educated at the « Oberschule » (non-classical secondary school) in Munich-Pasing and left after matriculating in 1948. After working for one year in industrial laboratories, he started reading physics at the Technical University (Technische Hochschule) in Munich in 1949 and passed his intermediate degree examinations in 1952. During the years 1953 and 1954 he completed his thesis at the Laboratory for Applied Physics at the Technical University in Munich, at the same time acting as assistant lecturer at its Institute of Mathematics. From 1955 to 1957 he worked on his thesis for the doctorate and carried out a series of investigations at the Institute for Physics of the Max Planck Institute for Medical Research in Heidelberg, in the course of which he carried out the first experimental observation of the phenomenon of Recoilless Nuclear Resonance Absorption. In January 1958 he received his degree under Professor Maier-Leibnitz at the Technical University in Munich. In 1958, again at the Max Planck Institute in Heidelberg, he provided the direct experimental evidence of the existence of Recoilless Nuclear Resonance Absorption. For the year 1959 he was appointed scientific assistant at the Technical University in Munich. He accepted an invitation by the California Institute of Technology in Pasadena, U.S.A., in 1960 and there continued his investigations of gamma absorption, at first as Research Fellow and later as Senior Research Fellow. He was appointed Professor of Physics at the California Institute of Technology in 1961.

From the year 1953 onwards his main work was directed towards the study of absorption of gamma rays in matter, in particular the study of nuclear resonance absorption. This led to the discovery of Recoilless Nuclear Resonance Absorption and its theoretical interpretation. During the last few years he has been investigating problems of nuclear physics and of solid state physics by applying already previously established methods.

His work in the field of Recoilless Nuclear Resonance Absorption has been rewarded by the following prizes: Prize of the Research Corporation

New York (1960); Röntgen Prize of the University of Giessen (1961); Elliot Cresson Medal, Franklin Institute (1961).

He is married to Elisabeth, *née* Pritz, and has two children, Peter and Regine.

Physics 1962

LEV DAVIDOVIČ LANDAU

«*for his pioneering theories for condensed matter, especially liquid helium*»

Physics 1962

Presentation Speech by Professor I. Waller, member of the Swedish Academy of Sciences

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

The winner of this year's Nobel Prize in Physics, Professor Lev Davidovič Landau at Moscow University, was born in Baku, 1908. His mathematical talents appeared at a very early age and at the age of 14 he began his studies at the University of Leningrad. After finishing them he spent one and a half years abroad, in particular with the well-known atomic physicist Niels Bohr in Copenhagen. He made a strong impression during this time thanks to his brilliant intellect and great outspokenness.

In 1930 Landau published a quantum theoretical investigation concerning the behaviour of free electrons in a magnetic field which immediately gave him international fame. This work turned out to be essential for the understanding of the properties of metals. Starting from new fruitful ideas Landau found after his return home, often in collaboration with his pupils, important results concerning the structure of magnetic substances and supraconductors and advanced fundamental theories for phase transformations and thermodynamical fluctuations.

Landau's ability to see the core of a problem and his unique physical intuition appear clearly in his investigations on liquid helium which he started after having been attached in 1937 to the Institute for Physical Problems in Moscow. The head of this institute was the famous physicist Kapitza who then performed interesting experiments on liquid helium. The natural helium gas had earlier been liquefied by cooling to about four degrees above the absolute zero of temperature and subsequent research had shown that this fluid when further cooled about two degrees was transformed to a new state which has quite strange properties. According to a term introduced by Kapitza it is superfluid which means that it can easily flow through very fine capillaries and slits which almost completely prevent the flow of all other liquids.

The originality in Landau's attack on the problem of explaining these phenomena was that he considered the quantized states of motion of the whole liquid instead of the states of the single atoms as other scientists had

done earlier. Landau started by considering the state of the fluid at the absolute zero temperature which is its ground state. He described the excited states of the liquid by the motion of certain fictive particles called quasi-particles. Landau combined experimental results with his calculations and deduced in this way the mechanical properties of the quasi-particles. These results, from which the properties of the fluid could be calculated, were later directly confirmed by investigations on the scattering of neutrons in liquid helium. Such experiments were first performed at Atomic Energy Ltd. in Stockholm in 1957. Landau further found that there exists in liquid helium besides ordinary sound waves also waves of a «second sound». He inspired thereby a Russian scientist to confirm this phenomenon experimentally.

Natural helium consists of an isotope of atomic weight 4 apart from about one millionth of another isotope of atomic weight 3. The lighter isotope has been studied in the liquid state since about 1950. This kind of liquid helium has properties which are quite different from those of the heavier isotope because the helium nuclei of atomic weights 3 and 4 are essentially different. A satisfactory theory for the lighter helium liquid was first given by Landau in 1956–1958 and has many formal similarities with his above-mentioned theory for the heavier isotope. The new theory is valid only at very low temperatures, less than one tenth of a degree above absolute zero. This is, however, the most interesting temperature range. Due to the difficulty of making measurements at these low temperatures the theory was not experimentally tested until very recently. These tests have been the more favourable for the theory the more the measuring technique has been refined. Landau has also predicted a new kind of wave propagation for this liquid and has called it zero sound. He has thereby stimulated experimental scientists to great efforts aiming to detect zero sound.

The importance of Landau's investigations are apparent when one considers that an important goal of physics research is to explain the properties of liquids as completely as it has been possible to explain the properties of crystals and of rarefied gases. In their efforts to attain this goal the scientists have in general met with insurmountable difficulties. An essential exception is Landau's theories of liquid helium which therefore are an achievement of great and profound importance.

Besides his investigations on condensed matter, i.e. matter in the solid and liquid state for which he is now awarded the Nobel Prize, Landau has also made contributions of the utmost importance to other parts of physics, in

particular to the theories of quantized fields and of elementary particles. He has by his original ideas and masterly investigations exercised far-reaching influence on the evolution of the atomic science of our time.

Professor Landau has unfortunately not yet fully recovered from the severe accident which he sustained at the beginning of this year. He is therefore not here to receive his Nobel Prize which is instead handed to him today by the Ambassador of Sweden in Moscow. On behalf of the Swedish Academy of Sciences I wish to express the hope that Professor Landau will soon completely recover.

No Lecture was delivered by Professor L. D. Landau

Biography

Lev Davidovič Landau was born in Baku on January 22, 1908, as the son of an engineer and a physician.

After graduating from the Physical Department of Leningrad University at the age of 19, he began his scientific career at the Leningrad Physico-Technical Institute. The years 1929–1931 he spent abroad, partly as a Rockefeller Foundation Fellow, working in Germany, Switzerland, England and, especially, in Copenhagen under Niels Bohr.

During 1932–1937 he was head of the Theoretical Department of the Ukrainian Physico-Technical Institute at Kharkov, and since 1937 he has been the head of the Theoretical Department of the Institute for Physical Problems of the Academy of Sciences of the U.S.S.R. in Moscow. Simultaneously he taught constantly as a professor of theoretical physics in the Kharkov and Moscow State Universities.

Landau's work covers all branches of theoretical physics, ranging from fluid mechanics to quantum field theory. A large portion of his papers refers to the theory of the condensed state. They started in 1936 with a formulation of a general thermodynamical theory of the phase transitions of the second order. After P. L. Kapitza's discovery, in 1938, of the superfluidity of liquid helium, Landau began extensive research which led him to the construction of the complete theory of the « quantum liquids » at very low temperatures. His papers of 1941–1947 are devoted to the theory of the quantum liquids of the « Bose type », to which the superfluid liquid helium (the usual isotope ^{4}He) refers. During 1956–1958 he formulated the theory of the quantum liquids of the « Fermi type », to which liquid helium of isotope ^{3}He refers.

In 1946 he was elected to the membership of the Academy of Sciences of the U.S.S.R. The U.S.S.R. State Prize was awarded to him several times, and in 1962 he received, jointly with E. M. Lifshitz, the Lenin Science Prize for their *Course of Theoretical Physics*.

Landau is a Foreign Member of the Royal Society (London), of the Danish Royal Academy of Sciences, of the Netherlands Royal Academy of Sciences, Foreign Associate of the National Academy of Sciences of the

U.S.A., Honorary Member of the American Academy of Arts and Sciences, of the Physical Society (London), and of the Physical Society of France. In 1961, he received the Max Planck Medal and the Fritz London Prize.

Name Index

Abbe, E. 237, 238, 241–243
Agnew, L. E. 510
Aigrain, P. R. R. 384
Alfvén, H. 479
Allen, H. S. 239
Alvarez, L. W. 209, 218, 518, 539, 546, 547
Amagat, E. H. 49, 50
Amaldi, E, E. 504, 515
Ambler, E. 399, 400
Anderson, C. D. 104, 105, 107, 111, 112, 114, 120, 121, 126, 220, 487, 490, 491, 505, 530
Anderson, Sir John 88
Andrews, T. 49
Appleton, Sir Edward V. 73–78, *79–89*
Arnold, H. D. 344
Aston, F. W. 193
Auger, P. 96, 114
Back, E. 28
Bainbridge, K. T. 171, 173
Baldo-Ceolin, M. 511
Balmer, J. J. 286
Baranger, M. 294
Bardeen, J. 61, 313–317, *318–343*, 345, 350, 376, 384, 385
Barnett, M. A. F. 75, 81
Barsukov, K. A. 467
Bay, Z. 275
Baeyer, H. J. von, *see* Von Baeyer, H. J.
Becker, G. E. 303
Becker, H. 274, 275, 278
Becker, J. A. 378
Benedicks, C. 320
Bergia, S. 577
Beringer, R. 308
Bethe, H. A. 107, 115, 132, 177, 263, 288, 293, 294
Bhabha, H. J. 107, 131
Bitter, F. 220
Bjorklund, R. 152
Blackett, P. M. S. 91–96, *97–122*
Blau, M. 103
Bloch, F. 197–202, *203–218*, 219, 221, 223, 226, 230, 579
Bloembergen, N. 228
Boddy, P. J. 385
Bohm, D. 478
Bohr, A. 221
Bohr, N. 6, 11, 14, 20, 25, 27–30, 38, 44, 203, 205, 217, 253–258, 260, 262, 265, 266, 271–274, 278, 284, 286, 299–305, 476, 560, 607, 611
Bolotovsky, B. M. 442
Bonhoeffer, K. F. 35
Born, M. 44, 61, 251–255, *256–270*, 278, 286, 561, 562
Bose, S. N. 33, 129, 130, 132, 611
Bothe, W. 104, 251–255, *271–279*
Bown, R. 341
Boyle, R. 49
Braddick, H. J. J. 114
Bradt, H. L. 146
Bragg, Sir Lawrence 121, 420
Brattain, W. H. 313–317, 318, 335–37, 340, 343, 350, 376, *377–386*
Breit, G. 82, 204, 284, 294, 302
Bridgman, P. W. 47–52, *53–72*, 342
Brode, R. B. 110, 115
Broglie, L. de, *see* De Broglie, L.
Brown, B. H. 377
Brown, J. 553
Brown, R. 154

Brown, W. 384
Brumberg, E. M. 428
Bryant, H. 553
Buckley, O. E. 344, 372
Bugg, D. V. 551
Builder, G. 82
Burch, C. R. 191
Burnstein, R. 553
Butler, C. C. 113–119, 121, 154, 530
Cailletet, L. 49
Carlson, A. G. 154
Čerenkov, P. A. 421–425, *426–441*, 442–444, 446, 447, 449, 451–453, 457–467, 472, 473, 475–478, 480, 483, 496–499, 505, 563, 583
Chadwick, Sir James 95, 103, 107, 109, 278
Chamberlain, O. 485–488, *489–507*, 509, 522
Champion, F. C. 103
Chen, N. Y. 387–392, 393, 397–400, *406–420*
Chew, G. F. 514, 515
Chudakow, A. E. 447
Clementel, E. 576–578
Cockcroft, Sir John D. 103, 158, 159, 161–166, *167–186*, 189, 193, 196
Compton, A. H. 254, 257, 271–273, 278, 426, 434, 582
Condon, E. U. 164, 167, 171
Cooper, L. N. 343
Corbino, O. M. 521
Cork, B. 509
Corson, D. R. 115, 521
Costa, J. L. 170
Coulomb, C. A. 130, 287, 294, 394, 476, 479, 533, 560, 561
Crane, H. R. 174
Cronin, J. 553
Curie, P. 426, 427
Curie-Sklodowska, M. 426, 427
Dalitz, R. 397
Dante, A. 508
Dayhoff, E. S. 290
Davisson, C. J. 13, 260, 261, 344, 375
Davydov, B. 319
De Broglie, L. 13, 30, 33, 36, 253, 256, 260–263, 286, 562
Debye, P. J. W. 217, 476, 594, 595, 600
Dee, P. I. 171, 173, 177
Dember 330, 335
Dennison, D. M. 34
Deutsch, M. 290
Dicke, R. H. 307
Dirac, P. A. M. 33, 40, 41, 95, 106, 120, 132, 208, 260, 263, 269, 283, 284, 286, 291, 294, 299, 305, 324, 390, 409, 410, 487, 489, 490, 505, 509, 519, 561, 571, 572, 575
Doppler, C. J. 449–453, 456, 457, 466, 467, 558, 586, 587, 596, 599
Dugas, C. 384
Dunoyer, L. 8
Durand, L. 577
Du Toit, A. L. 122
Ehrenfest, P. 60
Einstein, A. 27, 33, 106, 129, 130, 132, 164, 165, 254, 256–258, 262–265, 274, 292, 355, 424, 559, 592, 593
Ekspong, A. G. 504, 516, 518
Elliott, M. W. J. 447
Elsasser, W. 261
Epstein, P. S. 258
Estermann, I. 9, 14
Eucken, A. 35
Ewen, H. I. 201
Fabri, E. 397
Faraday, M. 78, 211, 223, 565
Faxén, H. 263
Fermi, E. 33, 126, 128, 132, 205, 217, 269, 324, 326, 327, 330, 331, 332, 334, 390, 404, 443, 475, 488, 506, 508, 518, 521, 569, 611
Fizeau, H. L. 9
Flügge, S. 551
Fock, V. 263
Foley, H. M. 302
Foucault, J. B. L. 481
Fowler, P. H. 146
Franck, J. 13, 120, 260, 262, 269

Frank, I. M. 421–425, 429, 431, 436, 441, *442–469*, 470, 473, 475, 483
Franklin, B. 315, 316
Fränz, H. 275, 278
Frazer, W. R. 576
Freier, P. 146
Frenkel, J. 319, 329, 330, 335, 336
Frisch, O. R. 551
Frisch, R. 14
Fulco, J. R. 576
Galilei, G. 393
Gamow, G. 164, 167, 171, 189
Gardner, J. H. 306, 308
Garibyan, G. M. 447
Garrett, C. G. B. 384, 385
Gauss, K. F. 200
Geiger, H. 93, 104, 144, 173, 189, 248, 261, 271–274, 277, 278
Gell-Mann, M. 512
Gerlach, W. 5, 11, 17
Germer, L. H. 13, 261
Gibbs, J. W. 33, 248
Gibney, R. B. 318, 336, 338
Gilbert, C. S. 533
Gilbert, W. 199
Ginzburg, V. L. 444, 445, 447, 448, 459, 470, 473, 474, 479, 480
Glaser, D. A. 523–528, *529–553*
Goeppert-Mayer, M. 269
Goldhaber, G. 504, 505, 515
Goldsmith, P. 447
Gorter, C. J. 200, 207, 212, 214, 228
Goucher, F. S. 367, 368
Goudsmit, S. 12, 30, 34, 286, 299
Graaff, R. J. van de, *see* Van de Graaff, R. J.
Graves, C. 553
Gross, E. P. 478
Gupta, S. 115
Gurney, R. W. 164, 167, 171
Györgyi, G. 474
Haga, H. 248
Hall 328, 330
Hamilton, J. 116
Hansen, W. W. 212, 218, 583
Hardy, T. C. 303
Harteck, P. 35, 173
Hartung, R. 553
Haynes, J. R. 321, 351, 353–355
Hayward, R. W. 399
Hazen, W. E. 114
Heald, M. A. 308
Heaviside, O. 75, 80, 81
Heisenberg, W. 14, 20, 30, 31, 32, 125, 128–130, 217, 253, 258–260, 262, 264, 269, 286, 287, 291, 586
Heitler, W. 107, 116
Henisch, H. K. 384
Hensby, G. S. 114
Herman, R. 577, 583
Hertz, G. 13, 262
Hertz, H. 78
Herzberg, G. 287
Hess, V. 104, 274
Hey, J. S. 88
Hilbert, D. 268
Hildebrand, R. H. 539
Hilsch, R. 320
Hipple, J. A. 222
Hoffman, G. 104
Hofstadter, R. 555–559, *560–583*
Holtsmark, J. 263
Hoppes, D. D. 399
Houston, W. V. 287
Hudson, E. P. 99
Hudson, R. P. 399
Hughes, D. J. 111
Hulthén, E. 5, 75, 199, 237, 487
Hund, F. 269
Hurwitz, A. 268
Huygens, C. 461, 463
Hylleraas, E. 269
Ising, G. 93, 188
Jánossy, L. 115, 116
Jelley, J. V. 438, 447
Jewett, F. B. 344
Joffe, A. 473
Joliot, F. 95, 165, 173
Joliot-Curie, I. 165, 173
Jones, H. 111

Jordan, P. 38, 39, 41, 258, 260, 269
Joule, J. P. 479
Kadyk, J. 553
Kakiuchi, Y. 225
Kapitza, P. 177, 185, 607, 611
Kapteyn, J. C. 248
Karplus, R. 309
Kelly, M. J. 341, 344
Kelvin, Lord 336, 381, 430
Kennedy, J. W. 521
Kennelly, A. E. 75, 80, 81
Kenrick, F. B. 533
Kepler, J. 27
Kerst, D. W. 188
Kingston, R. H. 384
Klein, O. B. 268, 389, 397
Koba, Z. 514
Koenig, S. H. 306
Kolhörster, W. 104, 255, 274, 278
Kolomensky, A. A. 467
Kramers, H. A. 217, 258, 271–274, 278, 284
Kroll, N. M. 294, 309
Krömer, H. A. 355
Kuhn, W. 584
Kunze, P. 104, 109
Kusch, P. 223, 281–285, 294, *298–312*
Ladenburg, R. 258
Lamb, H. 185
Lamb, W. E. 281–285, *286–297*, 591
Lambertson, G. R. 509
Landau, L. D. 478, 479, *605–612*
Landé, A. 25, 28
Langmuir, I. 10, 11, 378
Lao-tse 390
Laporte, O. 395, 396
Lark-Horovitz, K. 232, 320, 329
Larmor, J. 6, 20, 268, 444
Lattes, C. M. G. 141, 146
Lauritsen, C. C. 174
Lawrence, E. O. 165, 173, 188, 189, 488, 492, 504, 508
Lee, T. D. 387–392, 397–400, *406–420*
Lees, D. S. 99, 102, 103
Lenard, P. 277
Leprince-Ringuet, L. 114, 119
Lewis, G. N. 171, 173
Lewis, W. B. 175
Lifshitz, E. M. 611
Lindenbaum, S. 497, 498
Lindh, A. E. 49, 139
Lindström, G. 220
Lisell, E. 56
Little, J. B. 359
Livingston, M. S. 173, 188, 189
Lofgren, E. 504, 520
Loomis, F. W. 311
Lorentz, H. A. 248, 291, 394, 401, 411, 416, 449
Lovell, A. C. B. 113, 114, 121
Lummer, O. R. 242, 268
Lyot, B.-F. 245
Mach, E. 431, 472, 475
McIntyre, J. A. 582
McKay, K. G. 372
Mackenzie, K. R. 521
McMillan, E. M. 193, 488, 492, 504, 505
McSkimin, H. J. 321
Maier-Leibnitz, H. 273, 276, 559, 587, 602
Maiman, T. H. 290
Majorana, E. 129, 130
Mallet, M. L. 423, 427
Malmfors, K. G. 588
Mandelstam, L. I. 444, 453, 458, 480, 482
Mann, A. K. 305
Marconi, G. 75, 78, 79, 80
Marsden, E. 271
Marshak, R. E. 132
Marshall, J. 438
Marx, G. 474
Massey, Sir H. S. W. 263
Maxwell, J. C. 9, 10, 17, 37, 78
Mayer, J. 269
Meyer, D. 553
Michelson, A. A. 268, 286
Millikan, R. A. 104
Millman, S. 301, 302, 303

Minkowski, H. 268
Montén, F. G. A. 68
Moon, P. B. 587, 596, 597
Moore, H. R. 318
Morgan, S. O. 318
Morley, E. W. 286
Morozov, A. 480, 481
Morton, J. A. 359
Mössbauer, R. L. 555–559, 584–603
Mott, N. F. 103, 104, 263, 319, 561, 571
Moyer, B. J. 516
Muirhead, H. 141
Müller, J. 104, 120, 274
Nafe, J. E. 302
Nagle, D. 539
Naismith, R. 84
Nambu, Y. 576
Neddermeyer, S. H. 126
Nelson, E. B. 302
Newton, I. 256, 264, 393
Nicodemus, D. B. 213, 218
Nobel, A. 345
Nordheim, L. W. 111
Nuttall, J. M. 189
Occhialini, G. P. S. 94, 95, 103, 104, 105–107, 109, 112, 120, 141, 152
Ohl, R. S. 319, 320, 321, 337, 377
Ohm, G. S. 355, 357
Oliphant, M. L. E. 173, 177, 193
Oppenheimer, J. R. 269, 287, 292, 296, 419
Ornstein, L. S. 248
Packard, M. E. 212, 218
Pafomov, V. E. 447, 465
Paschen, F. 239
Pauli, W. 14, 20, 23–26, 27–45, 128, 181, 203, 217, 260, 269, 291, 489, 491, 560, 571, 572, 575
Pearson, G. L. 318, 327, 328, 350
Peng, H. W. 116
Perkins, D. H. 152
Perl, M. 553
Perrier, C. 521
Peters, B. 146
Peters, L. J. 342
Pevsner, A. 577
Pfann, W. G. 359, 360
Piccioni O. 146, 494, 504, 509, 519
Pickavance, T. G. 178
Pietenpol, W. J. 368
Piggott, W. R. 84
Pines, D. 477
Planck, M. 31, 33, 256, 257, 258, 260, 268, 271, 277, 286
Pohl, R. W. 320
Poisson, S. D. 362, 380, 534
Pose, H. 275
Pound, R. V. 201, 212, 226
Powell, C. F. 103, 112, 119, 126, 132, 137–143, *144–159*
Pringsheim, E. 268
Prins, J. A. 248
Proca, A. 37
Prodell, A. G. 306
Prowse, D. J. 511
Purcell, E. M. 197–202, 203, 210, 212, 218, *219–233,* 306, 308
Putten, J. van, *see* Van Putten, J.
Rabi, I. I. 3–7, 15, *19–21,* 200, 204, 207, 209, 232, 283, 284, 287, 298, 300, 302, 311
Rahm, D. 553
Raman, Sir C.V. 269
Rarita, W. 209
Ratcliffe, J. A. 82
Ravenhall, D. G. 579
Rayleigh, Lord 241
Regener, E. 104
Retherford, R. C. 288
Richards, R. E. 225
Richardson, J. R. 146
Richardson, Sir Owen 78
Ritz, W. 257
Rochester, G. D. 113–117, 121, 154, 530
Roe, B. 553
Roellig, L. 553
Röntgen, W. C. 437
Rosanes, J. 259, 268
Rosenbluth, M. 397, 572, 574, 575
Rosenfeld, A. H. 512

Rosenfeld, L. 38
Rosser, W. G. V. 119
Rossi, B. 104, 274, 275
Rowland, H. A. 239, 241
Rudberg, E. G. 315
Runcorn, S. K. 113
Runge, C. 239, 268
Russell, H. N. 303, 304
Rutherford, Lord Ernest 25, 87, 93, 97, 98, 101, 120, 121, 158, 163–165, 167, 169, 171, 173, 177, 185, 187, 189, 195, 262, 271, 557, 560, 561
Rydberg, J. R. 27
Ryder, E. J. 356, 357
Sagdeev, R. Z. 478
Sakata, S. 131
Sakurai, J. J. 577
Sanders, T. M. 290
Saunders, F. A. 303, 304
Scaff, J. H. 319, 320, 337, 376
Scheel, K. 248
Schenkel, M. 190
Scherrer, P. 217
Schiff, L. I. 579, 582
Schottky, W. 319, 330, 332
Schrieffer, J. R. 343
Schrödinger, E. 30, 217, 253, 256, 260–263, 274, 286
Schuster, Sir Arthur 122
Schwarzschild, K. 268
Schwinger, J. 209, 223, 285, 305, 306
Seaborg, G. T. 521
Segrè, E. G. 485–488, 489. 506, *508–522*
Serber, K. 188
Shafranov, V. D. 478
Shimizu, T. 97
Shive, J. N. 340, 350
Shockley, W. 313–317, 318, 320, 321, 335, 340, *344–376*, 379, 382
Siegbahn, K. 220, 423, 525, 565
Simon, Sir Francis 228
Simpson, O. C. 9
Sinclair, D. 553
Skinner, M. 290
Skobeltzyn, D. V. 104, 469
Skyrme, T. H. R. 179
Slater, J. C. 271–274, 375
Slätis, H. 551
Slepian, J. 187
Sloan, D. H. 189
Smaller, B. 220
Smith, J. A. S. 225
Soloviev, L. 481
Sommerfeld, A. 20, 27, 28, 44, 253, 258, 286, 287, 430, 473
Sparks, M. 358
Spenke 319
Statz, H. 384
Staub, H. H. 213, 218
Steiner, H. M. 504, 520
Stern, O. 3–7, 8–18, 20, 204, 203, 269, 521
Stewart, Balfour 79, 85
Stoner, E. C. 25, 29
Street, J. C. 112, 114
Takeda, G. 514
Tamm, I. E. 421–425, 436, 441, 459, *470–484*
Tanikawa, Y. 131
Tate, J. T. 311
Taub, H. 305
Taylor, H. S. 10
Teal, G. K. 359
Teller, E. 404
Theuerer, H. C. 319, 376
Thibaud, J. 95
Thomas, L. H. 30, 286
Thomson, G. P. 13, 154, 157, 261
Thomson, Sir J. J. 87, 154, 157, 268, 291
Tiomno, J. 397
Torrey, H. C. 212, 226
Triebwasser, S. 290
Trilling, G. 553
Tripp, R. D. 506
Tryggveson, O. 390
Tuve, M. A. 82
Tyndall, A. M. 158
Uehling, E. A. 294
Uhlenbeck, G. 12, 30, 34, 286, 299
Van de Graaff, R. J. 177, 193, 582

Van der Velde, J. 553
Van der Waals, J. D. 248
Van Putten, J. 553
Van Vleck, J. H. 229, 342
Vavilov, S. I. 423, 426, 428, 441, 442–444, 446, 447, 449, 451–453, 457–467, 469, *472*, 473, 475–478, 480, 483
Veksler, V. I. 193
Velde, J. van der, *see* Van der Velde, J.
Villi, C. 576–578
Vlasov, A. 478
Vleck, J. H. van, *see* Van Vleck, J. H.
Voigt, W. 268
Volmer, M. 533
Von Baeyer, H. J. 275
Von Weiszäcker, C. F. 112
Waals, J. D. van der, *see* Van der Waals, J. D.
Wagner, C. 329
Wahl, A. C. 521
Walker, M. 185
Waller, I. 25, 125, 163, 226, 229, 253, 283, 557, 594, 595, 600, 607
Walton, E. T. S. 103, 161–166, 167, 169, 171, 173, 185, *187–196*
Wambacher, H. 103
Wang, J. S. 404
Warburg, E. 271
Wataghin, G. 115
Watson-Watt, Sir Robert, 88
Webb, H. W. 288
Wegener, A. 122
Weisskopf, V. 221, 269, 292
Weiszäcker, C. F. von, *see* Von Weiszäcker, C. F.
Weizel, W. 232
Welton, T. A. 293
Wentzel, G. 262
Wenzel, W. A. 509
Weyl, H. 217, 415
Wheeler, J. A. 397
Wideröe, R. 187, 189
Wieder, I. 290
Wiegand, C. 489, 504, 506, 520, 522
Wiener, E. 571
Wiener, N. 262
Wigner, E. P. 38, 39, 41, 128, 342, 396
Williams, E. J. 111, 112, 115
Williams, R. C. 287
Wilson, A. H. 319, 323, 329
Wilson, C. T. R. 91, 93, 94, 97, 98, 139, 140, 145, 158, 159, 261, 272, 525, 526
Wilson, H. A. 122, 319, 323, 329
Wilson, J. G. 112, 113, 114
Wilson, P. 511
Wimett, T. F. 220, 221
Winckler, J. 438
Wismer, K. L. 533
Wittke, J. B. 307
Wood, J. G. 539
Wu, C.S. 391, 399, 400
Wu, T. Y. 404
Yager, W. A. 321
Yang, C. N. 387–392, *393–405*, 406, 419, 420
Yearian, M. R. 575
Yennie, D. R. 579
Ypsilantis, T. 489, 504, 506, 520, 522
Yukawa, H. 123–127, *128–136*, 152, 514, 557, 577
Zeeman, P. 11, 14, 28–30, 34, 203–205, 289, 290, 299, 521
Zener, C. 372
Zernike, F. 235–238, *239–249*
Zhdanov, A, 139
Zhelesniakov, V. V. 479
Zipf, T. 553

Subject Index

Antinucleons 508
Antiprotons 489
Atomic nuclei, interaction with high-speed nucleons 167
Bubble chambers, elementary particles and 529
Cloud-chamber researches 97
Coincidence method 271
Cosmic radiation 144
Cosmic radiation, cloud-chamber researches in 97
Electron, magnetic moment of 298
Electron-scattering method 560
Elementary particles and bubble chambers 529
Exclusion principle 27
Fast particles, artificial production of 187
Fine structure of hydrogen atom 286
Gamma radiation, recoilless nuclear resonance absorption of 584
High-pressure physics 53
High-speed nucleons, interaction with atomic nuclei 167
Hydrogen atom, fine structure of 286
Induction, nuclear 203
Interactions, weak 406
Ionosphere 79
Law of parity conservation 393
Light sources moving in refractive media, optics of 442
Magnetic moment of electron 298
Magnetism, nuclear 219
Meson theory 128
Molecular rays 8
Nuclear induction 203
Nuclear magnetism 219
Nuclear physics, cloud chamber researches in 97
Nuclei, structure of 560
Nucleons, high-speed, interaction with atomic nuclei 167
Nucleons, structure of 560
Parity, conservation of 393
Parity, nonconservation of 406
Particles (see also pertinent entries)
Particles, elementary, and bubble chambers 529
Particles, fast, artificial production of 187
Particles, moving at super-light velocities 426, 470
Phase contrast 239
Plasma physics 470
Point-contact transistor 318
Quantum mechanics 27
Quantum mechanics, statistical interpretation of 256
Radiation of particles moving at super-light velocities 426, 470
Rays, molecular 8
Recoilless nuclear resonance absorption of gamma radiation 584
Semiconductor research 318
Semiconductors, surface properties of 377
Super-light velocities, particles moving at 426, 470
Symmetry laws of physics 393
Transistor, point-contact 318
Transistor technology 344

Index of Biographies

Appleton, E. V. 87
Bardeen, J 342
Blackett, P. M. S. 120
Bloch, F. 217
Born, M. 268
Bothe, W. 277
Brattain, W. H. 385
Bridgman, P. W. 71
Čerenkov, P. A. 441
Chamberlain, O. 506
Cockcroft, J. D. 185
Frank, I. M. 469
Glaser, D. A. 552
Hofstadter, R. 582
Kusch, P. 311
Lamb, W. E. 296
Landau, L. D. 611
Lee, T. D. 419
Mössbauer, R. L. 602
Pauli, W. 44
Powell, C. F. 158
Purcell, E. M. 232
Rabi, I. I. 20
Segrè, E. G. 521
Shockley, W. 375
Stern, O. 17
Tamm, I. E. 483
Walton, E. T. S. 195
Yang, C. N. 404
Yukawa, H. 135
Zernike, F. 247