Cambridge University Press
978-0-521-33837-0 - The Birth of Particle Physics
Edited by Laurie M. Brown and Lillian Hoddeson
Excerpt
More information

Part I

Introduction

1 The birth of elementary particle physics: 1930–1950

LAURIE M. BROWN AND LILLIAN HODDESON

In the 1930s and 1940s, physicists significantly revised their views on the elementary constituents of matter, which during the 1920s had been widely assumed to be only the electron and the proton. This revision and the problems it posed gave birth to the field of modern elementary particle physics. The Fermilab symposium on which this volume is based was held to illuminate the study of this historical process through lectures by important participants and through discussions among physicists and historians. We believe that these chapters and the supplementary material that was collected after the symposium provide a useful examination of the origins of particle physics.

However, this volume is not a definitive or even comprehensive history of those origins. The lack of prior historical work in this area provided both a challenge and a rationale, but at the same time a handicap; during and after the symposium we became aware of important omissions. Some gaps were filled with postsymposium contributions, but others remain, as will be pointed out in this chapter. In addition, much of the material consisted of personal recollections, valuable to the historian as source material, but not balanced and critical historical accounts. People cannot be totally objective about the events in which they participate; we tend unconsciously to reinterpret history in terms of present-day values. Thus, expert witnesses may, in subtle ways, undercut the program of the historian to understand the past in its own terms.

Other problems stem from professional differences between physicists and historians of science. Physicists must keep their eyes on the cutting edge of current research; they aim to advance the frontier and, if possible, to do so before anyone else. This requires formulating general concepts applicable to a large class of phenomena – abstract models that exclude all but the most relevant attributes of real objects and processes. But historians of science are concerned with identifying and characterizing the sequence of earlier frontiers and with explaining, within the context of their period, how each frontier led to the next. To do this, the historian must examine, in each period, and in rich complexity, the particular circumstances that influenced a given development. Such differences of focus can interfere with communication between physicists and historians.[1] We were, indeed, in inviting physicists to deliver the major historical lectures at the symposium, asking them to step out of character and view particle physics from an alien point of view. It is remarkable that they not only accepted this almost impossible assignment but did extremely well at it – a success we attribute, in part, to the presence of historians in the audience.

The professional differences between the two groups did produce tensions at the symposium (notably in the two round-table discussions, Chapters 16 and 17) that were generally constructive. They generated debate and called forth efforts to be more precise and to document their statements. It is through this tension that this volume differs from other books of recollections.[2] In retrospect, we believe that we should have gone further in this direction. We suggest that in future symposia of this kind, historians, as well as physicists, be invited to present papers to the mixed group.

Besides relating episodes in the adventure of twentieth century physics and serving as a source of data and references, this volume is a collection of viewpoints that may suggest working hypotheses for future work on the history of particle physics. One such hypothesis, used as a guide in selecting symposium speakers, was that particle physics emerged out of the turbulent confluence of three initially distinct bodies of research: nuclear physics, cosmic-ray studies, and quantum field theory. By the mid-1930s, where these fields overlapped there was conflict and apparent paradox, partly resolved by the end of the 1940s (although the resolution posed new and urgent problems). Because an earlier symposium held at the University of Minnesota in Minneapolis-St. Paul in 1977 had already considered many of the relevant nuclear physics issues,[3] we concentrated on those origins of particle physics derived from cosmic-ray studies and quantum field theory. To help the reader find a suitable vantage point from which to assess the individual contributions, we shall attempt in this introductory chapter to sketch the broad landscape of particle physics in the 1930s

Cambridge University Press
978-0-521-33837-0 - The Birth of Particle Physics
Edited by Laurie M. Brown and Lillian Hoddeson
Excerpt
More information

and 1940s: The first section will characterize the three bodies of research out of which particle physics emerged during the 1930s; the second section will deal with a problem not adequately discussed at the symposium, the application of a suspect quantum electrodynamics (QED) to the analysis of cosmic-ray events; the third section will review important experiments and theories that implied the existence of new particles; the remaining three sections will place the results in a broader framework.

The sources of particle physics

By 1930, relativity and quantum mechanics were established, and it was clear that matter is made of nuclear atoms. But the excitement of the new physics was far from over. Indeed, the next half century was characterized by startling experimental and theoretical discoveries and an assortment of puzzles wherever one looked – in particular, in the three currents that flowed together to make particle physics.

The atomic nucleus[4]

The consensus of scientists in 1930 was that matter was made of two elementary particles, electrons and protons, negative and positive electricity.[5] The neutral atom of mass number A and charge number Z was believed to contain A protons and $A - Z$ electrons in its nucleus and Z electrons in its outer shells. All forces relevant on the atomic scale were believed to be electromagnetic.

Some hoped that the tiny nucleus, containing most of the mass but only a million-millionth part of the volume of the atom, could itself be treated as a quantum mechanical system.[6] Other voices argued that just as a new dynamic theory (quantum mechanics) was needed to pass from the "human scale" to the atom, so a new generalization of the physical laws might be needed at the microcosmic distances in the nucleus.[7] Indeed, since early in the century it had been predicted (notably by Hendrik Antoon Lorentz) that the electromagnetic theory of Maxwell would fail at distances smaller than the "classical electron radius" ($r_0 = e^2/mc^2 \sim 2.8 \times 10^{-13}$ cm, where e and m are the charge and mass of the electron and c is the speed of light), which happens to be approximately the nuclear radius.

The difficulties encountered in treating the nucleus as a quantum mechanical system of protons and electrons were these: The nucleus was supposed to contain A protons and $A - Z$ electrons, but when the latter number is odd, as for ^{6}Li and ^{14}N, the spin and statistics (i.e., whether Bose or Fermi type) are incorrect. Also, unpaired electron spins in the nucleus would imply hyperfine splitting of atomic spectral lines on a scale about 1,000-fold larger than is observed. Another nuclear puzzle was that it was impossible (in the relativistic quantum theory of the electron) to confine the light electron within the small nucleus. Finally, there was the continuous spectrum of β-decay electron energies, which called into question even the conservation of energy.

Radical suggestions were seriously considered for modifying the mechanics and/or the electrodynamics and even the conservation laws. But the resolution was to hinge on the existence of new particles: the neutron, discovered by James Chadwick in 1932, and the neutrino, proposed in 1930 by Wolfgang Pauli and incorporated in a theory of β decay by Enrico Fermi in 1934.[8] These two neutral particles permitted the banishment of electrons from nuclear models. The positron, discovered in the cosmic rays in 1932 by Carl David Anderson, was used to complete the picture of nuclear β decay when Irène Curie and Frédéric Joliot produced artificially radioactive light elements that decayed by positron emission.

The cosmic rays

Cosmic rays were discovered (Table 1.1) as a result of investigations, after 1900, of fine-weather "atmospheric electricity" (i.e., in the absence of an electrical thunderstorm). At about this time it was realized that the nonzero electrical conductivity of the atmosphere implied the presence of an ionizing agent. Indeed, minute amounts of radioactive substances, first discovered in 1896 by Henri Becquerel, were found to be present in the earth, air, and water (and in the measuring instruments themselves), and these did provide ionizing rays. However, after all known sources of ionization had been accounted for, there remained a "residual" conductivity, even in closed vessels that were heavily shielded. This phenomenon implied the existence of a "penetrating radiation" of unknown origin.[9]

The altitude dependence of the atmospheric conductivity was investigated by means of balloon flights, mainly in central Europe, and notably by Victor F. Hess in Austria. The manned balloons carried sealed electrometers whose rates of discharge first decreased with altitude, but then (above 2 km) began a marked increase. This pattern of ionization suggested, if it did not prove, the existence of an extrater-

Table 1.1 *Sequence of development of cosmic-ray physics*[a]

I. Prehistory (to 1911, especially from 1900)

"Atmospheric electricity" during calm weather; conductivity of air measured by electrometers; connection with radioactivity of earth and atmosphere; interest was also geophysical and meteorological.

II. Discovery (1911-14) and exploration (1922-30)

Balloons carrying observers with electrometers measured the altitude dependence of ionization and showed that there is an ionizing radiation that comes from above; these measurements began in 1909 and continued (at intervals) to about 1930, in the atmosphere, under water, earth, etc.; the primaries were assumed to be high-energy photons from outer space; search for diurnal and annual intensity variations; study of energy inhomogeneity.

III. Particle physics, early (1930–47)

Direct observation of the primaries was not yet possible, but "latitude effect" showed that they were charged particles; secondary charged-particle trajectories were observed with cloud chambers and counter telescope arrays, and momentum was measured by curvature of trajectory in a magnetic field; discovery of positron and of pair production; soft and penetrating components; radiation processes and electromagnetic cascades; meson theory of nuclear forces; discovery of mesotron (present-day muon); properties of the muon, including mass, lifetime, and penetrability; two-meson theory and the meson "paradox."

IV. Particle physics, later (1947-53)

Observation of particle tracks in photographic emulsion; discovery of pion and pion-muon-electron decay chain; nuclear capture of negative pions; observation of cosmic-ray primary protons and fast nuclei; extensive air showers; discovery of the strange particles; the strangeness quantum number.

V. Astrophysics (1954 and later)

Even now the highest-energy particles are in cosmic rays, but such particles are rare; studies made with rockets and earth satellites; primary energy spectrum, isotopic composition; x-ray and γ-ray astronomy; galactic and extragalactic magnetic fields.

[a] In successive periods, at least one change occurred that was so significant that it required a totally new interpretation of the previous observations and theories.

restrial source for the penetrating radiation, so that eventually (by the late 1920s) one spoke of the *cosmic rays* (Table 1.1). Until 1930, their specific ionization (ions per cubic centimeter per second) was the only property systematically observed. The vertical variation of the rate of discharge of an electrometer was interpreted as measuring an absorption length of the rays in air or water and showed them to have increasing penetrability (hardening) with depth.[10] At that time, the main interest in the rays was geophysical and astrophysical, although Robert A. Millikan proposed a theory of cosmic-ray origins that involved a new physical mechanism.[11]

The focus changed at the end of the 1920s, when two methods, the cloud chamber with magnetic field and coincidence counting, were used to study the individual behavior of charged particles produced by collisions of the primaries with air molecules. Both were adapted from techniques used to study x rays and radioactivity. They were flexible, permitting a variety of experiments to be performed, and they could be combined. Their descendants are the principal tools used today to study the interactions of elementary particles, whether the source be cosmic rays or accelerators. The pioneers in this enterprise were Walther Bothe and Werner Kolhörster in Berlin and Dmitry Skobeltzyn in Leningrad.

In 1924, Bothe and Hans Geiger had measured the degree of simultaneity of emission of the recoil electron and the scattered x-ray photon in Compton scattering, in order to test a statistical theory proposed by N. Bohr, H. A. Kramers, and J. C. Slater.[12] The Bothe-Geiger apparatus used two Geiger point counters, each connected to a separate string electrometer. Counts were registered by deflection of the string, and both electrometers were photographed by a camera with motor-driven film transport. By the time they finished the experiment, Bothe and Geiger had achieved a time resolution of 10^{-4} sec.

Kolhörster, a colleague of Bothe at the Physikalisch-Technische Reichsanstalt in Charlottenburg, outside Berlin, and an experienced cosmic-ray worker, pointed out in 1928 that by aligning two point counters in a vertical array, Bothe's coincidence technique could be used to make a γ-ray telescope for cosmic rays.[13] A few days later, Bothe and Kolhörster reported similar observations with the far more efficient Geiger-Müller tube counter. By mid-1929 they established that a 4.1-cm gold block placed between the counters reduced the coincidence rate by only 24%, and they concluded from this that the primary rays had "corpuscular nature."[14] Until then they had been thought to be high-energy photons and had been called (e.g., by Hess) "ultra γ rays."

To young Bruno Rossi, at the physics laboratory of the University of

Florence in Arcetri, Italy, these results reported from Berlin "came like a flash of light revealing the existence of an unsuspected world," and he began immediately the exploration of this world.[15] He soon found a way to improve the technique using a vacuum tube circuit to detect the coincident discharges of the tube counters, achieving greater flexibility and time resolution, and by using three out-of-line counters, he discovered that there was a great abundance of secondary radiation, later identified as "cascade showers."

Meanwhile, in Leningrad, Skobeltzyn, who had been studying radioactive γ radiation, began using the Wilson cloud chamber to observe the trajectories of cosmic-ray particles in a magnetic field, where a charged particle's track is curved, with a radius of curvature directly proportional to the particle momentum and inversely proportional to the magnetic field. He also noted that tracks appeared to be associated with each other, to a degree difficult to account for by the scattering processes known at that time.[16] Skobeltzyn's work was the first visual method for observing processes of particles with energies higher than those available from radioactive sources.

Skobeltzyn's counterpart in California was Carl David Anderson, who had been using a cloud chamber to study photoelectrons produced by x rays, and who wanted to move on to Compton collisions of nuclear γ rays; instead, at the urging of his boss, Millikan, he began tooling up (in 1930-1) a cloud chamber and a strong magnetic field to observe cosmic-ray interactions. Anderson was to discover two new particles in cosmic rays: the positron and the muon.[17]

The other major step forward was the invention and use in 1932 of the counter-controlled cloud chamber by P. M. S. Blackett and G. P. S. Occhialini.[18] In such a chamber, both the expansion and camera are activated by an electronic pulse from a counter array that selects a class of events, so that, in effect, the incident particle "takes its own picture." Soon after Anderson had discovered what he referred to as "easily deflectable positives," Blackett and Occhialini used their new instrument to observe electron pair production (suggested by P. A. M. Dirac's electron theory) and cascade showers. In the words of Hanson, "they discovered that the 'Anderson particle' and the 'Dirac particle' were the same particle."[19] By 1930, therefore, the technical framework had been established for the spectacular cosmic-ray discoveries of the next two decades (including new particles) made using counter and cloud chamber techniques.

Relativistic spinning electrons and quantum fields

The prediction of the positron, which began not as a triumph but as an embarrassment, produced a profound alteration in our notion of "particle." In his first article on the new electron theory, Dirac stated that he looked for "some incompleteness" in the Heisenberg and Schrödinger treatments of the point electron "such that when removed, the whole of the duplexity phenomena [would] follow without arbitrary assumptions."[20] He was referring to the spinning-electron model of Samuel Goudsmit and George Uhlenbeck, and Pauli's use of that model to account for details of atomic spectra by a doubling of allowed states.[21] The incompleteness was found to be the lack of relativistic invariance, which he corrected by proposing a new wave equation consistent with his transformation theory.[22] An important result was that the electron was forced to have spin of one-half and magnetic moment of one Bohr magneton. Furthermore, in the new theory, the state of the electron was described by four wave functions, rather than one as in Schrödinger's theory, or two as in Pauli's nonrelativistic theory of the spinning electron. There were thus two new degrees of freedom related to the existence of states of both signs of the energy.

The embarrassment came from Einstein's theory of relativity, in which, according to the relation $E = mc^2$, negative energy always means negative mass. This was a meaningless concept, as C. G. Darwin pointed out.[23] Dirac could not evade this issue. Much later he said:

> The problem of the negative-energy states puzzled me for quite a while. The main method of attack to begin with was to try to find some way of avoiding the transitions to negative-energy states, but then I approached the question from a different point of view. I was reconciled to the fact that the negative-energy states could not be excluded from the mathematical theory, and so I thought, let us try and find a physical explanation for them.[24]

He was thus led to consider a world in which all, or nearly all, of the infinite number of negative energy states were already occupied by an electron – "filled," if we take account of Pauli's exclusion principle, which prevents more than a single electron from occupying a given state. Transitions from a positive to a negative energy state would practically not occur. There remained the problem of the physical interpretation of the occasional unoccupied state, or "hole," in the sea of negative energy electrons. After a period of doubt, the hole was recognized to be the positive electron; and it became fully legitimate in 1932 when it was experimentally observed.[25]

The theory of holes was based on "a very close analogy provided by the chemical theory of valency," in which an unfilled hole in an otherwise closed atomic shell behaves like a positively charged valence electron.[24] In the case of the negative-energy states, there is no positive background charge to neutralize the vacuum; also, one must deal not with a finite number of electrons but with an infinite number of electrons. The last entails an infinite number of quantum mechanical degrees of freedom and thus, in turn, quantum fields.

Relativistic electron theory and the quantum theory of fields were both on the agenda of theoretical physics after the invention of quantum mechanics by Heisenberg and Schrödinger in 1925 and 1926. Both emerged from the fertile brain of Paul Dirac, once he stopped being "strongly fascinated by Bohr orbits," as he says in Chapter 2, and formulated his general quantum Hamiltonian dynamics. In his pioneering work of February 1927 on quantum electrodynamics (QED), Dirac proposed a solution to the problem of the wave-particle duality, which had puzzled physicists since Einstein hypothesized the light quantum in 1905.[26] At the end of this paper, Dirac summarized its contents as follows:

> The problem is treated of an assembly of similar systems satisfying the Einstein-Bose statistical mechanics, which interact with another different system, a Hamiltonian function being obtained to describe the motion. The theory is applied to the interaction of an assembly of light-quanta with an ordinary atom, and it is shown that it gives Einstein's laws for the emission and absorption of radiation.
>
> The interaction of an atom with electromagnetic waves is then considered, and it is shown that if one takes the energies and phases of the waves to be q-numbers satisfying the proper quantum conditions instead of c-numbers, the Hamiltonian function takes the same form as in the light-quantum treatment. The theory leads to the correct expressions for Einstein's A's and B's.

(The As and Bs are light-quantum emission and absorption probability amplitudes.) From this we can see that Dirac treated the electromag-