Physics 1984

CARLO RUBBIA and SIMON VAN DER MEER

for their decisive contributions to the large project, which led to the discovery of the field particles W and Z, communicators of weak interaction

THE-NOBEL PRIZE FOR PHYSICS

Speech by Professor GÖSTA EKSPONG of the Royal Academy of Sciences. Translation from the Swedish text

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen,

This year's Nobel Prize for Physics has been awarded to Professor CARLO RUBBIA and Dr. SIMON VAN DER MEER. According to the decision of the Royal Swedish Academy of Sciences the prize is given "for their decisive contributions to the large project, which led to the discovery of the field particles W and Z, communicators of the weak interaction".

The large project mentioned in the citation is the antiproton project at CERN, the international centre for research devoted to the study of elementary particles, which has 13 European states as members. CERN straddles, in a unique way, the border between two countries, Switzerland and France, and has grown progressively in importance over the 30 years of its life. The international character is underlined by the fact that Carlo Rubbia is Italian, Simon van der Meer is Dutch and the collaborators in the various phases of the project are scientists, engineers, and technicians of many nationalities, either employed by CERN or in one of the many universities or research institutes involved in the experiments. The project has been made possible by collaboration, by the pooling of financial resources and of scientific and technical skill. When the antiproton project was proposed eight years ago the CERN ship had two captains-two Directors General, Professor Leon van Hove from Belgium and Sir John Adams, from the United Kingdom. Navigating through the high waves generated by the convincing enthusiasm of Rubbia but having van der Meer on board as pilot to steer through the more difficult waters, they directed their ship towards new challenging frontiers. The late Sir John Adams had been responsible for the construction of the two outstanding proton accelerators, which were called into action in new roles for the new project.

A former Nobel Laureate expressed his opinion about the CERN project with the following words: van der Meer made it possible, Rubbia made it happen. Looking closer one finds that two conditions had to be fulfilled in order to produce the W and Z in particle collisions: The first is that the particles must collide at sufficiently high energy so that the conversion of energy into mass could create the heavy W and Z particles. The second is that the number of collisions must be large enough to give a chance of seeing the rare creation process taking place. The name of Rubbia is connected with the first condition, that of van der Meer with the second. Rubbia's proposal was to use the largest accelerator at CERN, the SPS, as a storage ring for circulating antiprotons as well as for protons, circulating in the opposite direction. The particles in the two beams would cross the French-Swiss border about 100,000 times every second for a whole day or more, to be repeated with new beams during months of operation. Antiprotons cannot be found in nature, in any case not on Earth.

But they can be created at CERN where sufficient energy is available at the other accelerator, the PS. The antiprotons are accumulated in a special storage ring, built by a team led by van der Meer.

It is here that his ingenious method, called stochastic cooling, enables an intense antiproton beam to be built up. The signals from produced particles are recorded in huge detector systems set up around two collision points along the periphery of the SPS storage ring. The largest of these detectors was designed, built and operated by a team led by Rubbia. A second large detector was built by another team, operating it in parallel with the first one, nicely confirming the extremely important results.

An old dream was fulfilled last year when the discoveries of the W and Z were made at CERN-the dream of better understanding the weak interaction, which turns out to be weak just because the W and Z are so very heavy. The weak interaction is unique in that it can change the nature of a particle, for example transforming a neutron into a proton or vice versa. Such transformations are crucial for the sun and it is the weakness of the interaction which leads to the very slow burning of the nuclear fuel of the sun and thus creates the conditions on earth which can support life.

At first radioactive decays were the only weak interaction phenomenon available for study. Nowadays thanks to accelerators and storage rings this field of research is quite large. The theory which synthesizes a vast amount of knowledge and combines our understanding of the weak and electromagnetic interactions was honoured by the award of the Nobel Prize for physics in 1979 to Sheldon Glashow, Abdus Salam and Steven Weinberg. It also predicted new phenomena caused by the invented particle Z, introduced to make the theory consistent. Such phenomena were first observed in a CERN experiment about ten years ago. The only historical parallel goes back 120 years to Maxwells theory for electric and magnetic phenomena. In that case the theory was made consistent by a new ingredient, which contained the seed for the prediction of radio waves, discovered by Heinrich Herz almost 100 years ago. The modern electroweak theory contains not only the electromagnetic photons as communicators of force but also the communicators W and Z which act as a kind of shock-absorber, especially noticeable in hard collisions-such as those which must have occurred frequently during the Big Bang era at the early stage of the evolution of our universe. The collisions in the CERN collider may be hard enough to break loose the communicators, the shock-absorbers, for a short moment. The resulting fireworks of newly produced particles have been observed in the detectors, and the signs showing the presence of the W and Z have been seen and a start has been made on measuring their properties.

Professor Rubbia and Dr. van der Meer,

Your achievements in recent years, leading to the successful operation of the CERN proton-antiproton collider, have been widely admired in the whole world. The discovery of the W and Z particles will go down in the history of physics like the discovery of radio waves and the photons of light, the communicators of electromagnetism.

I know that you share your-joy with many collaborators at CERN and in the participating universities. I also know that they congratulate you in many ways, also by setting new records for energy and for the rate of collisions, and by finding new interesting phenomena produced in the collisions. The discovery of the W and Z is not the end-it is the beginning.

On behalf of the Royal Swedish Academy of Sciences, I have the pleasure and -the honour of extending to you our warmest congratulations. I now invite you to receive your prizes from the hands of His Majesty the King.

Carlo Rubbia

CARLO RUBBIA

I was born in the small town of Gorizia, Italy, on 31 March, 1934. My father was an electrical engineer at the local telephone company and my mother an elementary school teacher. At the end of the World War II most of the province of Gorizia was overtaken by Yugoslavia and my family fled to Venice first and then to Udine.

As a boy, I was deeply interested in scientific ideas, electrical and mechanical, and I read almost everything I could find on the subject. I was attracted more by the hardware and construction aspects than by the scientific issues. At that time I could not decide if science or technology were more relevant for me.

After completing High School, I applied to the Faculty of Physics at the rather exclusive Scuola Normale in Pisa. My previous education had been seriously affected by the disasters of the war and the subsequent unrest. I badly failed the admission tests and my application was turned down. I forgot about physics and I started engineering at the University of Milan (Politecnico). To my great surprise and joy a few months later I was offered the possiblity of entering the Scuola Normale. One of the people who had won the admission contest had resigned! I am recollecting this apparently insignificant fact since it has determined and almost completely by accident my career of physicist. I moved to Pisa, where I completed the University education with at thesis on cosmic ray experiments. They have been very tough years, since I had to greatly improve my education, which was very deficient in a number of fundamental disciplines. At that time I also participated under my thesis advisor Marcello Conversi to new instrumentation developments and to the realization of the first pulsed gas particle detectors.

Soon after my degree, in 1958 I went to the United States to enlarge my experience and to familiarize myself with particle accelerators. I spent about one and a half years at Columbia University. Together with W. Baker, we measured at the Nevis Syncro-cyclotron the angular assymmetry in the capture of polarized muons, demonstrating the presence of parity violation in this fundamental process. This was his first of a long series of experiments on Weak Interactions, which ever since has become my main field of interest. Of course at that time it would have been quite unthinkable for me to imagine to be one day amongst the people discovering the quanta of the weak field!

Around 1960 I moved back to Europe, attracted by the newly founded European Organization for Nuclear Research, where for the first time the idea of a joint European effort in a field of pure Science was to be tried in practice. The Syncro-cyclotron at CERN had a performance significantly superior to the one of the machine in Nevis and we succeeded in a number of very exciting experiments on the structure of weak interactions, amongst which I would like to mention the discovery of the beta decay process of the positive pion, $\pi^+ = \pi^0 + e + \nu$ and the first observation of the muon capture by free hydrogen, $\mu^- + p = n + \nu$.

In the early sixties John Adams brought to operation the CERN Proton-Syncrotron. I moved to the larger machine where I continued to do some weak interaction experiments, like for instance the determination of the parity violation in the beta decay of the lambda hyperon.

During the Summer of 1964 Fitch and Cronin announced the discovery of CP violation. This has been for me a tremendously important result and I abandoned all current work to start a long *series* of observations on CP-violation in K^0 decay and on the K_L - K_S mass difference. Unfortunately the subject did not turn out to be as prolific as in the case of the previous discovery of parity violation and even today, some thirty years afterwards we do not know much more about the origin of CP-violation than right after the announcement of the discovery.

I returned again to more orthodox weak interactions a few years later, when together with David Cline and Alfred Mann we proposed a major neutrino experiment at the newly started US laboratory of Fermilab. The operational problems associated with a limping accelerator and a new laboratory made very difficult, albeit impossible for us during the Summer of 1973 to settle definitively the question of the existence of neutral currents in neutrino interactions, when competing with the much more advanced instrumentation of Gargamelle at CERN. Instead, about one year later we could cleanly observe the presence of di-muons events in neutrino interactions and to confirm in this way one of the crucial predictions of the GIM mechanism, hinting at the existence of charm, glamorously settled only few months later with the observation of the Ψ/J particle.

In the meantime and under the impulse of Vicky Weisskopf a new, fascinating adventure had just started at CERN with a new type of colliding beams machine, the Intersecting Storage Rings, in which counter-rotating beams of protons collide against each other. This novel technique offered a much more efficient usc of the accelerator energy than the traditional method of collisions against a fixed target. From the very first operation of this new type of accelerator, I have participated to a long series of experiments. They have been crucial to perfect the detection techniques with colliding beams of protons and antiprotons needed later on for the discovery of the Intermediate Bosons.

By that time it was quite clear that Unified Theories of the type SU (2) *x* U (1) had a very good chance of predicting the existence and the masses of the triplet of intermediate vector bosons. The problem of course was the one of finding a practical way of discovering them. To achieve energies high enough to create the intermediate vector bosons (roughly 100 times as heavy as the proton) together with David Cline and Peter Mc Intyre we proposed in 1976 a radically new approach. Along the lines discussed about ten years earlier by the russian physicist Budker, we suggested to transform an existing high energy accelerator in a colliding beam device in which a beam of protons and of antiprotons, their antimatter twins, are counter-rotating and colliding head-on. To this effect we had to develop a number of techniques for creating antiprotons, confining them in a concentrated beam and colliding them with an intense proton beam. These techniques were developed at CERN with the help of many people and in particular of Guido Petrucci, Jacques Gareyte and Simon van der Meer.

In view of the size and of the complexity of the detector, physics experiments at the proton-antiproton collider have required rather unusual techniques. Equally unusual has been the number and variety of different talents needed to reach the goal of observing the W and Z particlcs. International cooperation between many people from very different countries has been proven to be a very successful way of achieving such goals.

(added in 1991) : For eighteen years, I have dedicated one semester per year to teaching at Harvard University in Cambridge, Mass., where I have been appointed professor in 1970, spending the rest of my time mostly in Geneva, where I was conducting various experiments, especially the UA-1 Collaboration at the proton-antiproton collider until 1988.

On 17 December 1987, the Council of CERN decided to appoint me Director-General of the Organization as from 1st January 1989, for a mandate of five years.

My wife, Marisa, teaches Physics at High School, and we have two children, a married daughter Laura, medical doctor, and a son, Andre, student in high energy physics.

EXPERIMENTAL OBSERVATION OF THE INTERMEDIATE VECTOR BOSONS W^+, W^- and Z^0.

Nobel lecture, 8 December, 1984

by

CARLO RUBBIA

CERN, CH-1211 GENEVA 23, Switzerland

1. *Introduction*

In this lecture I shall describe the discovery of the triplet of elementary particles W^+, W^-, and Z^0-by far the most massive elementary particles produced with accelerators up to now. They are also believed to be the propagators of the weak interaction phenomena.

On a cosmological scale, weak interactions play an absolutely fundamental role. For example, it is the weak process

$$p + p \rightarrow {}^2H + e^+ + \nu_e$$

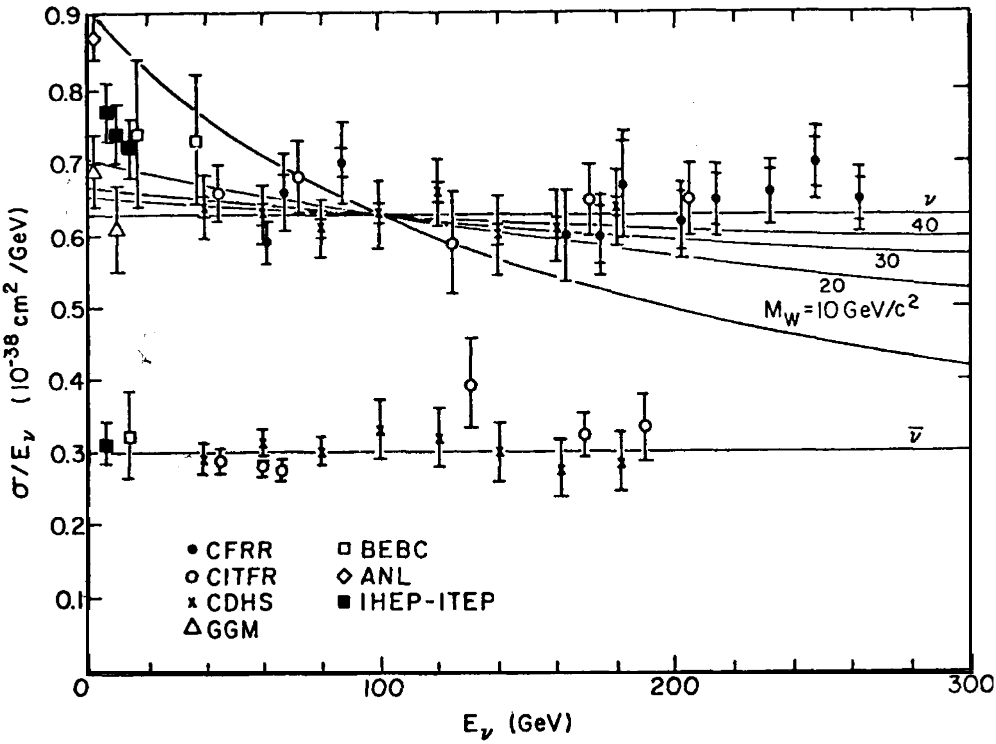
that controls the main burning reactions in the sun. The most striking feature of these phenomena is their small rate of occurrence: at the temperature and density at the centre of the sun, this burning process produces a heat release per unit of mass which is only 1/100 that of the natural metabolism of the human body. It is indeed this slowness that makes them so precious, ensuring, for instance, the appropriate thermal conditions that are necessary for life on earth. This property is directly related to the very large mass of the W-field quanta.

Since the fundamental discoveries of Henri Becquerel and of Pierre and Marie Curie at the end of the last century, a large number of beta-decay phenomena have been observed in nuclei. They all appear to be related to a pair of fundamental reactions involving transformations between protons and neutrons:

$$n \rightarrow p + e^- + \nu_e, \qquad p \rightarrow n + e^+ + \nu_e. \tag{1}$$

Following Fermi [1], these processes can be described perturbatively as a point interaction involving the product of the four participating fields.

High-energy collisions have led to the observation of many hundreds of new hadronic particle states. These new particles, which are generally unstable, appear to be just as fundamental as the neutron and the proton. Most of these new particle states exhibit weak interaction properties which are similar to those of the nucleons. The spectroscopy of these states can be described with the help of fundamental, point-like, spin-1/2 fermions, the quarks, with fractional electric charges +2/3e and -1/3e and three different colour states. The universality of the weak phenomena is then well interpreted as a Fermi



Fig, I. The muon neutrino and antineutrino charged-current total cross-section as a function of the neutrino energy. Data are from the Particle Data Group (Rev. Mod. Phys. 56, No. 2, Part 2, April 1984) reprinted at CERN. The lines represent the effects of the W propagator.

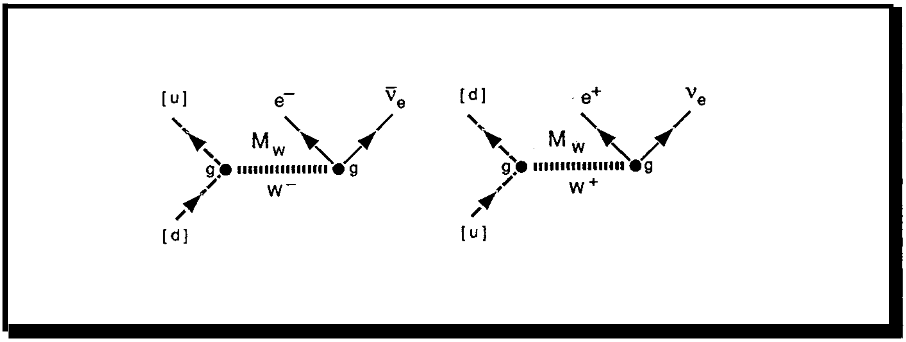
coupling occurring at the quark level [2]. For instance, reactions (1) are actually due to the processes

$$(d)\rightarrow(u)+e^-+\tilde{\nu}_e, \qquad (u)\rightarrow(d)+e^++\nu_e, \tag{2}$$

where (u) is a +2/3e quark and (d) a -l/3e quark. (The brackets indicate that particles are bound.) Cabibbo has shown that universality of the weak coupling to the quark families is well understood, assuming that significant mixing occurs in the +1/3e quark states [3]. Likewise, the three leptonic families -namely (e, ν_e), (μ, ν_μ), and (τ, ν_τ)-exhibit identical weak interaction behaviour, once the differences in masses are taken into account. It is not known if, in analogy to the Cabibbo phenomenon, mixing occurs also amongst the neutrino states (neutrino oscillations).

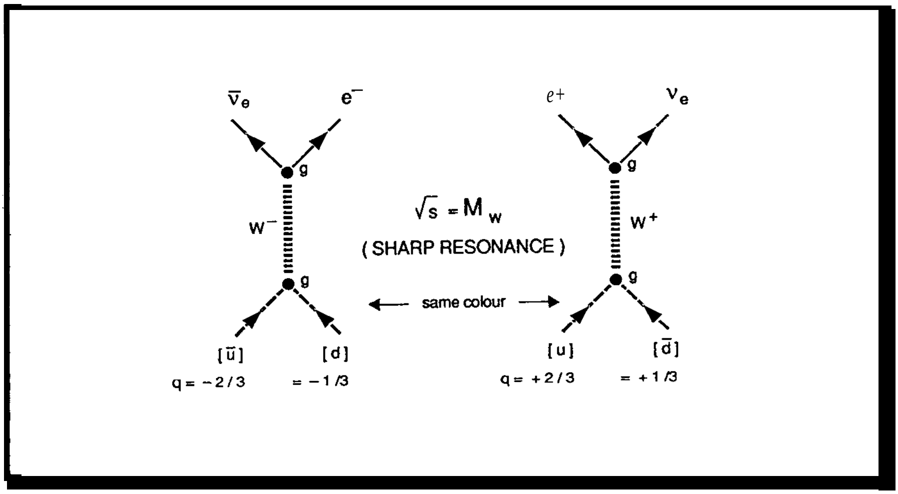
This has led to a very simple perturbative model in which there are three quark currents, built up from the (u, d_C), (c, s_C), and (t, b_C) pairs (the subscript C indicates Cabibbo mixing), and three lepton currents from (e, ν_e), (μ, ν_μ), and (τ, ν_τ) pairs. Each of these currents has the standard vector form [4] $J_\mu=f_1\gamma_\mu(1-\gamma_5)f_2$. Any of the pair products of currents J_μ, j_μ, will relate to a basic four-fermion interaction occurring at a strength determined by the universal Fermi constant G_F:

$$\mathscr{L}(x)=(G_F/\sqrt{2})J_\mu^*(x)j^\mu(x)+\text{c.c.},$$



a)

Fig. 2a. Feynman diagram of virtual W exchange mediating the weak process [reaction (2)]



b)

Fig. *2b*. Feynman diagram for the direct production of a W particle. Note that the quark transformation has been replaced by a quark-antiquark annihilation.

where $G_F = 1.16632 \times 10^{-5}\,\mathrm{GeV}^{-2}$ $(\hbar = c = 1)$.

This perturbative, point-like description of weak processes is in excellent agreement with experiments, up to the highest q^2 experiments performed with the high-energy neutrino beams (Fig. 1). We know, however, that such a perturbative calculation is incomplete and unsatisfactory. According to quantum mechanics, all higher-order terms must also be included: they appear, however, as quadratically divergent. Furthermore, at centre-of-mass energies greater than about 300 GeV, the first-order cross-section violates conservation of probability.

It was Oskar Klein [5] who, in 1938, first suggested that the weak interactions could be mediated by massive, charged fields. Although he made use of Yukawa's idea of constructing a short-range force with the help of massive field quanta, Klein's theory established also a close connection between electromagnetism and weak interactions. We now know that his premonitory vision is embodied in the electroweak theory of Glashow, Weinberg and Salam [6], which will be discussed in detail later in this lecture. It is worth quoting Klein's view directly:

'The role of these particles, and their properties, being similar to those of the photons, we may perhaps call them "electro-photons" (namely electrically charged photons).'

In the present lecture I shall follow today's prevalent notation of W^+ and W^- for these particles-from 'weak' [7]--although one must recognize that Klein's definition is now much more pertinent.

The basic Feynman diagrams of reaction (2) are the ones shown in Fig. 2a.

The new, dimensionless coupling constant g is then introduced, related to $G_F/\sqrt{2} \cong g^2/m_W^2$, for $q^2 << m_W^2$. The V-A nature of the Fermi interaction requires that the spin J of the W particle be 1. It is worth remarking that in Klein's paper, in analogy to the photon, J= 1 and *g=a*. The apparently excellent tit of the neutrino data to the four-fermion point-like interaction (Fig. 1) indicates that mw is very large (560 GeV/c^2) and is compatible with $m_W = \infty$.

2. *Production of W particles*

Direct production of W particles followed by their decay into the electron-neutrino is shown in Fig. 2b. The centreof-mass energy in the quark-antiquark collision must be large enough, namely $\sqrt{s} \simeq m_W$. The cross-section around the resonance will follow a characteristic Breit-Wigner shape, reminiscent of nuclear physics experiments. The cross-section is easily calculated:

$$\sigma(q\bar{q} \rightarrow W) = \tfrac{3}{4}\pi\lambda^2\Gamma_i\Gamma/[(E-m_W)^2+\Gamma^2/4],$$

where λ is the reduced quark wavelength in the centre of mass. Quark and antiquark must have identical colours. The initial-state width $\Gamma_i \equiv \Gamma_{q\bar{q}} \simeq$ $4.5\times10^{-7}\,m^3$(GeV) calculated from G_F is surprisingly wide: namely, for $m_W \simeq 82$ GeV/c^2 as predicted by SU(2) x U(1) theory, $\Gamma_{q\bar{q}} \cong 450$ MeV. The total width Γ depends on the number of quark and lepton generations. Taking $N_q=3$ and $N_\ell=3$, again for $m_W \simeq 100$ GeV, we find $\Gamma = 4\times\Gamma_{q\bar{q}} = 2$ GeV.

At the peak of the resonance,

$$\sigma(q\bar{q} \rightarrow W, \sqrt{s} = m_W) = 3\pi\lambda_i^2 B_i,$$

where $B_i = \Gamma_i/\Gamma$ is the branching ratio for the incoming channel.

Of course quark-antiquark collisions cannot be realized directly since free quarks are not available. The closest substitute is to use collisions between protons and antiprotons. The fraction of nucleon momentum carried by the quarks and antiquarks in a proton is shown in Fig. 3. Because of the presence of antiquarks, proton-proton collisions also can be efficiently used to produce W particles. However, a significantly greater beam energy is needed and there is no way of identifying the directions of the incoming quark and antiquark. As we shall see, this ambiguity will prevent the observation of important asymmetries associated with parity (P) and charge (C) violation of weak interactions. The centre-of-mass energy in the quark-antiquark collision $s_{q\bar{q}}$ is related to $S_{p\bar{p}}$ by the well-known formula,

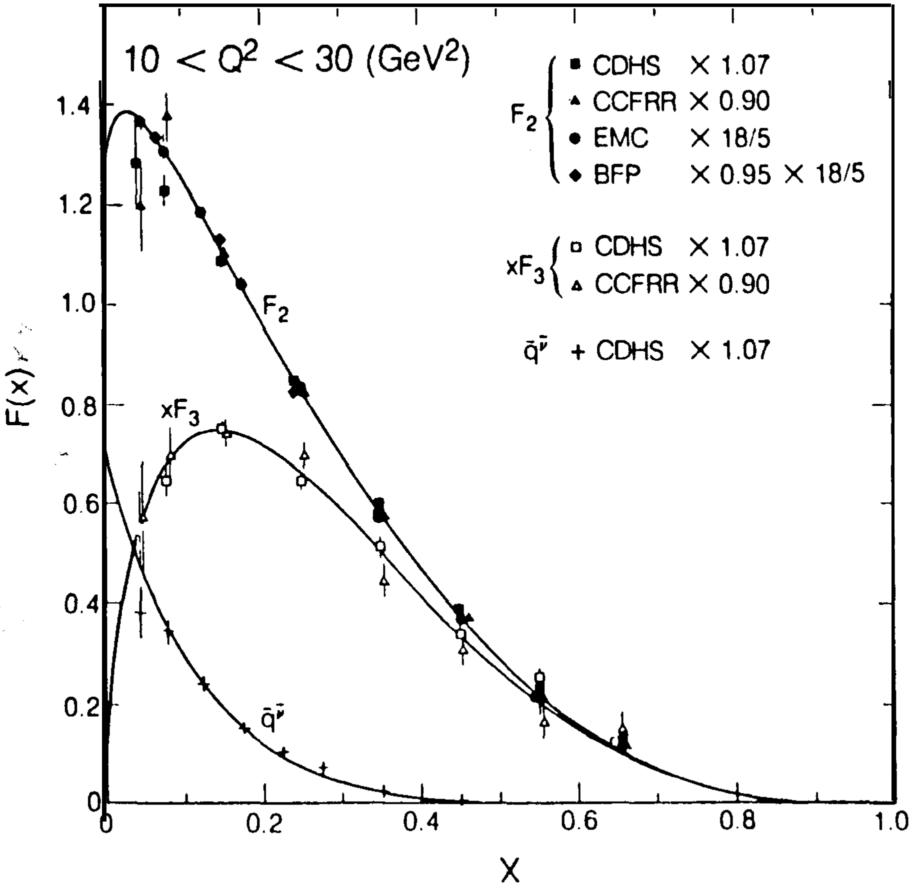


Fig. 3. Structure functions F_2, xF_3, and $\bar{q}^{\bar{\nu}}$, measured in different experiments, for fixed Q^2 versus x, plotted assuming $R=\sigma_L/\sigma_T=0$. The electromagnetic structure function $F_2^{\mu N}$ measured by the EMC (European Muon Collaboration) and the BFP [Berkeley (LBL) - FNAL- Princeton] is compared with the charged-current structure function $F_2^{\nu N}$ using the 18/5 factor from the average charge squared of the quarks. No correction has been applied for the difference between the strange and charm sea quarks, so the interpretation is $F_2 = x[q + \bar{q} - 3/5(s + \bar{s} - c - \bar{c})]$. (In this Q^2 range, $F_2^{\nu N}$ is depleted by a similar amount due to charm threshold effects in the transition $s \rightarrow c$.) The antiquark distribution measured from antineutrino scattering is $\bar{q}^{\bar{\nu}}=x(\bar{u}+\bar{d}+2\bar{s})$. The solid lines have the forms: $F_2=3.9x^{0.55}(1-x)^{3.2}+1.1(1-x)^8$, $xF_3=3.6x^{0.55}(1-x)^{3.2}$, $\bar{q}^{\bar{\nu}}=0.7(1-x)^8$. Relative normalization factors have been fitted to optimize agreement between the different data sets, and absolute changes have been arbitrarily chosen as indicated. [References: CDHS--H. Abramowicz et al., Z. Phys. *C17*, 283 (1983); CCFRR-F. Sciulli, private communication; EMC-J.J. Aubert et al., Phys. Lett. *105B*, 322 (1981); and A. Edwards, private communication; BFP--A. R. Clark et al., Phys. Rev. Lett. 51, 1826 (1983); and P. Meyers, Ph. D. Thesis, LBL-17108 (1983), Univ. of Calif., Berkeley. Courtesy J. Carr, LBL.]

$$s_{q\bar{q}} = S_{p\bar{p}} \cdot x_p \cdot x_{\bar{p}}.$$

Note that according to Fig. 3, in order to ensure the correct correlation between the quark of the proton (and the antiquark of the antiproton) the energy should be such that $x_p \simeq x_{\bar{p}} \gtrsim 0.25$ Therefore there is one broad optimum

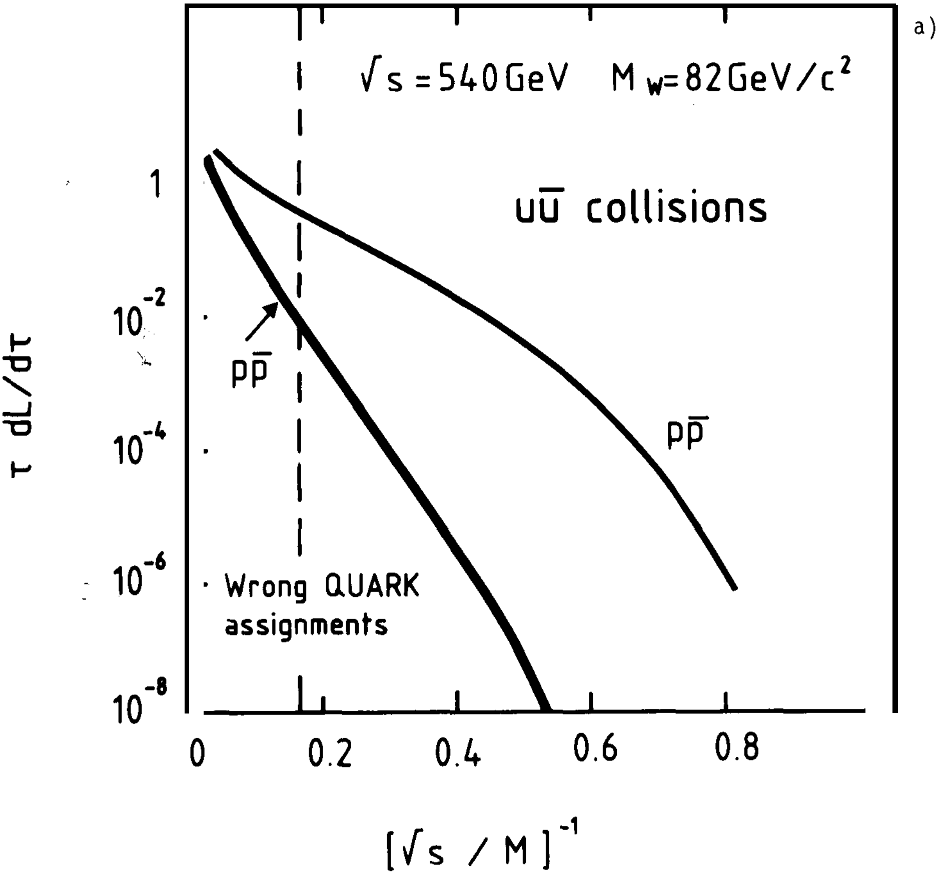


Fig. 4a, b. Production cross-sections of intermediate vector bosons for proton-antiproton collisions. The mass is parametrized with $\tau^{-1/2} = \sqrt{s}/M$. Note in Fig. 4a the small probability of wrong quark-antiquark assignments. The prints in Fig. 4b relate to mass predictions for the SU(2) x U(1) model.

energy range for the proton-antiproton collisions for a given W mass. For $m_w = 80\,\mathrm{GeV}/c^2$, $\sqrt{S_{p\bar{p}}} \simeq 400$–$600$ GeV. The production cross-section for the process

$$p\bar{p} \rightarrow W^{\pm} + X, \quad W^{\pm} \rightarrow e^{\pm} + \nu_e$$

(where X denotes the fragmentation of spectator partons) can be easily evaluated by folding the narrow resonance width over the p and $\bar{p}$ momentum distributions (Fig. 4). For $m_w = 82$ GeV/c and $\sqrt{S_{p\bar{p}}} = 540$ GeV, one finds $\sigma \cdot B = 0.54 \times 10^{-33}\,\mathrm{cm}^2$.

3. Proton-antiproton collisions

The only practical way of achieving centre-of-mass energies of the order of 500 GeV is to collide beams of protons and antiprotons [8]. For a long time such an idea had been considered as unpractical because of the low density of beams when used as targets.

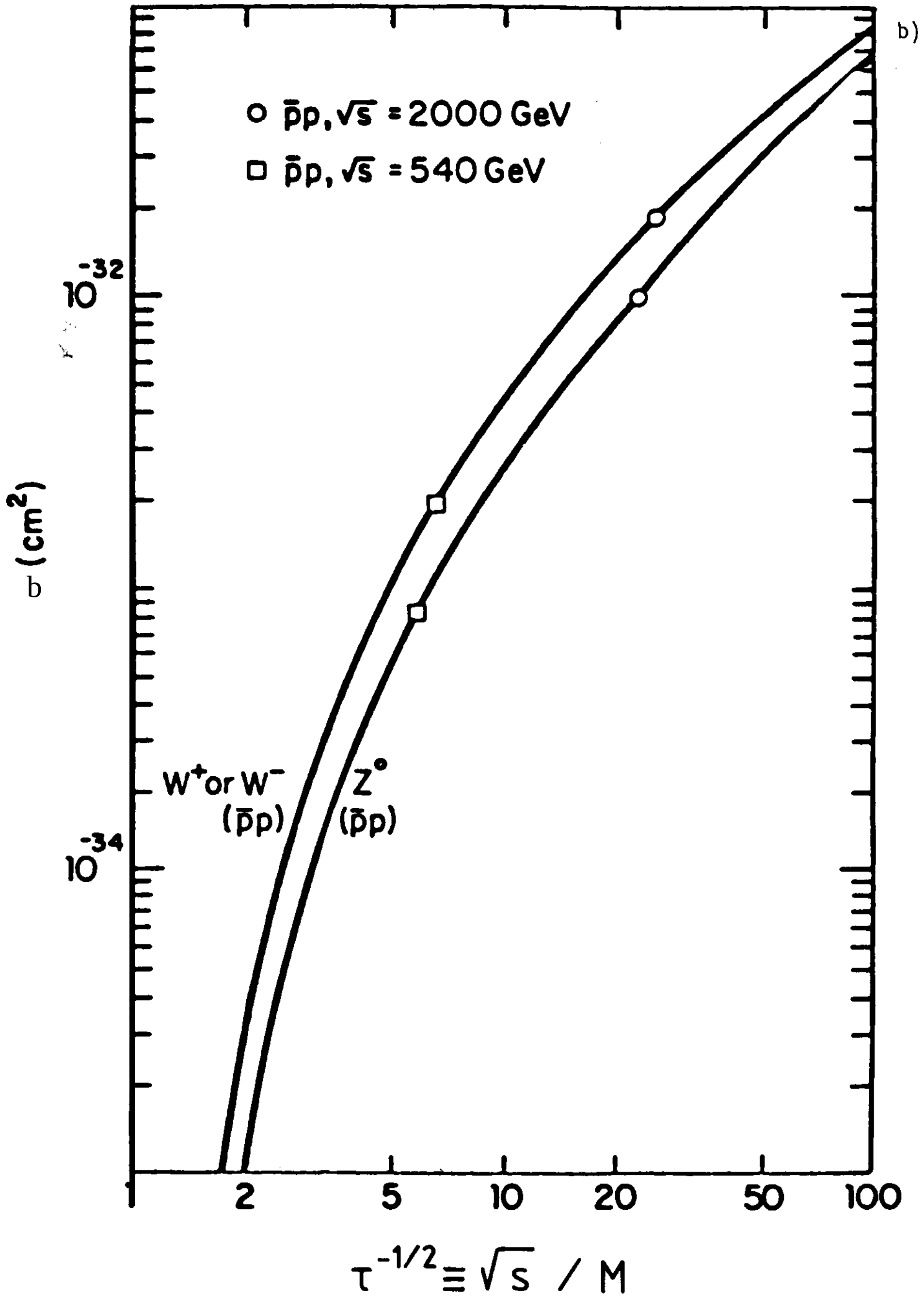

b)
○ $\bar{p}p, \sqrt{s} = 2000$ GeV
□ $\bar{p}p, \sqrt{s} = 540$ GeV
10^{-32}
10^{-34}
σ (cm^2)
W^+ or W^-
($\bar{p}p$)
Z^0
($\bar{p}p$)
1
2
5
10
20
50
100
$\tau^{-1/2} \equiv \sqrt{s} / M$

Fig. 4

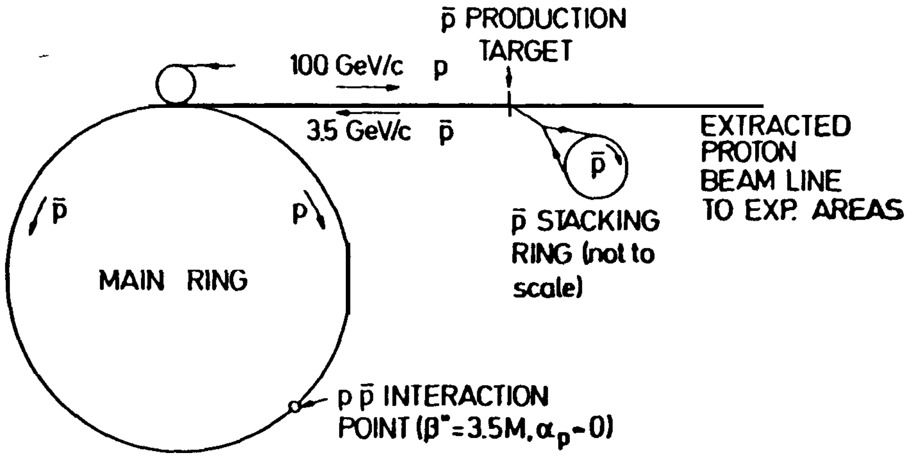


Fig. 5. **General layout of the $p\bar{p}$ colliding scheme**, from Ref. [9]. Protons (100 GeV/c) are periodically extracted in short bursts and produce 3.5 GeV/c antiprotons, which are accumulated and cooled in the small stacking ring. Then $\bar{p}$'s are reinjected in an RF bucket of the main ring and accelerated to top energy. They collide head on against a bunch filled with protons of equal energy and rotating in the opposite direction.

The rate R of events of cross-section σ for two counter-rotating beam bunches colliding head on, with frequency f_0 and n_1 and n_2 particles, is

$$R = (n_1 n_2 f_0/4)(\sigma/\pi\varrho^2),$$

where ϱ is the (common) beam radius, and the numerical factor 1/4 takes into account the integration over Gaussian profiles. For our experiment, typically $\varrho = 0.01$ cm and $\sigma = 10^{-34}$ cm^2. Therefore $(\sigma/\pi\varrho^2) = 3 \times 10^{-31}$, and a very large $n_1 n_2$ product is needed to overcome the 'geometry' effect.

The scheme used in the present experimental programme has been discussed by Rubbia, Cline and McIntyre [9] and is shown in Fig. 5. It makes use of the existing 400 GeV CERN Proton Synchrotron (PS) [10], suitably modified in order to be able to store counter-rotating bunches of protons and antiprotons at an energy of 270 GeV per beam. Antiprotons are produced by collisions of 26 GeV/c protons from the PS onto a solid target. Accumulation in a small 3.5 GeV/c storage ring is followed by stochastic cooling [ll] to compress phase space. In Table 1 the parameters of Ref. [9] are given. Taking into account that the original proposal was formulated for another machine, namely the Fermilab synchrotron (Batavia, Ill.) they are quite close to the conditions realised in the SPS conversion. Details of the accumulation of antiprotons are described in the accompanying lecture by Simon van der Meer.

The CERN experiments with proton-antiproton collisions have been the first, and so far the only, example of using a storage ring in which bunched protons and antiprotons collide head on. Although the CERN $p\bar{p}$ Collider uses *bunched* beams, as do the e^+e^- colliders, the phase-space damping due to synchrotron radiation is now absent. Furthermore, since antiprotons are

Table 1. List of parameters (from Ref. [9])

1. MAIN RING (Fermilab)	
- Beam momentum	*250 (400)* GeV/c
- Equivalent laboratory energy for ($p\bar{p}$)	*133 (341)* TeV
- Accelerating and bunching frequency	*53.14* MHz
- Harmonic number	*1113*
- RF peak voltage/turn	3.3×10^{6} v
- Residual gas pressure	$< 0.5 \times 10^{-7}$ Torr
- Beta functions at interaction point	*3.5* m
- Momentum compaction at int. point	-0 m
- Invariant emittances ($N_p = 10^{12}$)	
- longitudinal	3 eV.s
- transverse	$50\ \pi\ 10^{-6}$ rad.m
- Bunch length	*2.3* m
- Design luminosity	5×10^{29} (8 x 10") $cm^{-2}s^{-1}$
2. ANTIPROTON SOURCE (Stochastic Cooling [11]	
- Nominal stored $\bar{p}$ momentum	*3.5* GeV/c
- Circumference of ring	100 m
- Momentum acceptance	*0.02*
- Betatron acceptances	$100\ \pi\ 10^{-6}$ rad.m
- Bandwidth of momentum stochastic cooling	*400* MHz
- Maximum stochastic accelerating RF voltage	*3000* v
- Bandwidth of betatron stochastic cooling	*200* MHz
Final invariant emittances ($N_p = 3 \times 10^{10}$)	
- longitudinal	0.5 eV.s
- transverse	$10\ \pi\ 10^{-6}$ rad.m

scarce, one has to operate the collider in conditions of relatively large beam-beam interactions, which is not the case for the continuous proton beams of the previously operated Intersecting Storage Rings (ISR) at CERN [12]. One of the most remarkable results of the $p\bar{p}$ Collider has probably been the fact that it has operated at such high luminosity, which in turn means a large beam-beam tune shift. In the early days of construction, very serious concern had been voiced regarding the instability of the beams due to beam-beam interaction. The beam-beam force can be approximated as a periodic succession of extremely non-linear potential kicks. It is expected to excite a continuum of resonances of the storage ring which has, in principle, the density of rational numbers. Reduced to bare essentials, we can consider the case of a weak antiproton beam colliding head on with a strongly bunched proton beam. The increment, due to the angular kick Ax', of the action invariant W= $\gamma x^2 + 2\alpha x x' + \beta x'^2$ of an antiproton is $\Delta W = \beta(\Delta x') + 2(\alpha x + \beta x')\Delta x'$, and this can be expressed in terms of the 'tune shift', AQ as $\Delta x' = 4\pi\Delta Q x/\beta$. If we now assume that the successive kicks are randomized, the second term of AW averages to zero, and we get

$$\langle \Delta W/W \rangle = \tfrac{1}{2}(4\pi\Delta Q)^2.$$

For the design luminosity we need AQ-0.003, leading to $(\Delta W/W) = 7.1 \times 10^{-4}$. This is a very large number indeed, giving an e-fold in-

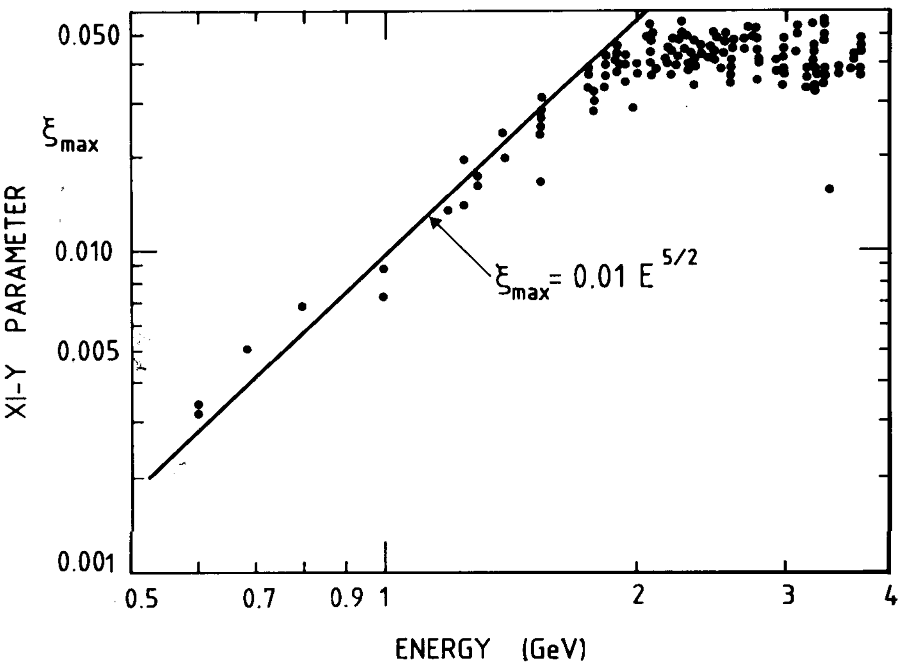


Fig. 6. Maximum allowed beam-beam tune-shift parameter, XI-Y, as a function of energy of the electron-positron collider SPEAR. One can see a dramatic drop in the allowed tune shift at lower energies, as a consequence of the reduced synchrotron damping. Extrapolation to the case of proton-antiproton collisions where the damping is absent and therefore the damping time is constant, is to be identified with the beam lifetime, permitting an infinitesimal tune shift and therefore to an unpractical luminosity.

crease of W in only $1/7.1 \times 10^{-4} = 1.41 \times 10^{3}$ kicks! Therefore the only reason why the antiproton motion remains stable is because these strong kicks are not random but periodic, and the beam has a long 'memory' which allows them to be added coherently rather than at random. Off-resonance, the effects of these kicks then cancel on the average, giving an overall zero amplitude growth. The beam-beam effects are very difficult, if not impossible, to evaluate theoretically, since this *a priori* purely deterministic problem can exhibit stochastic behaviour and irreversible diffusion-like characteristics.

A measurement at the electron-positron collider SPEAR at Stanford had further aggravated the general concern about the viability of the $p\bar{p}$ collider scheme. Reducing the energy of the electron collider (Fig. 6) resulted in a smaller value of the maximum allowed tune shift, interpreted as being due to the reduced synchrotron radiation damping. Equating the needed beam lifetime for the $p\bar{p}$ collider (where damping is absent) with the extrapolated damping time of an e^+e^- collider gives a maximum allowed tune shift $\Delta Q = 10^{-5} \div 10^{-6}$, which is catastrophically low. This bleak prediction was not confirmed by the experience at the collider, where $\Delta Q = 0.003$ per crossing, and six crossings are routinely achieved with a beam luminosity lifetime approaching one day. What, then, is the reason for such a striking contradiction between experiments with protons and those with electrons? The difference is caused by the presence of synchrotron radiation in the latter case. The emission of synchrotron photons is a major source of quick randomization between crossings and leads to a rapid deterioration of the beam emittance. Fortunately, the same phenomenon also provides us with an effective damping mechanism. The $p\bar{p}$ collider works because both the randomizing and the damping mechanisms are absent. This unusually favourable combination of effects has ensured that $p\bar{p}$ colliders have become viable devices. They have the potential for substantial improvements in the future. The accumulation of more antiprotons would permit us to obtain a substantially larger luminosity, and a project is under way at CERN which is expected to be able to deliver enough antiprotons to accumulate, *in one single* day, the integrated luminosity on which the results presented in this lecture have been based (~ 100 nb^{-1}).

4. The detection method

The process we want to observe is the one represented in Fig. 2b, namely

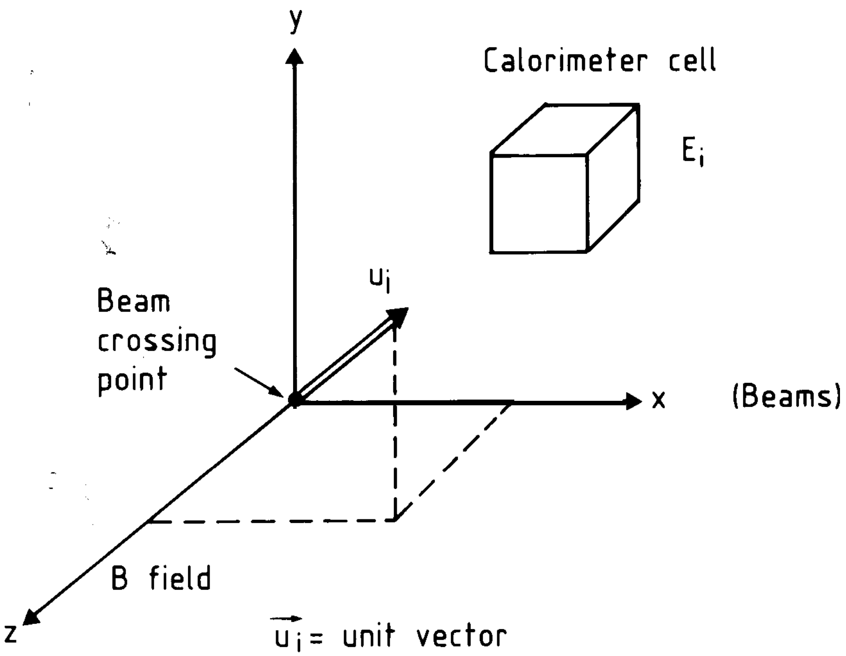
$$p+\bar{p}\rightarrow W^{\pm}+X\ ,\quad W^{\pm}\ e^{\pm}+\ \nu_e\ , \tag{3}$$

where X represents the sum of the debris from the interactions of the other protons (spectators). Although the detection of high-energy electrons is relatively straightforward, the observation of neutrino emission is uncommon in colliding-beam experiments. The probability of secondary interactions of the neutrino in any conceivable apparatus is infinitesimal. We must therefore rely on kinematics in order to signal its emission indirectly. This is achieved with an appropriately designed detector [13] which is uniformly sensitive, over the whole solid angle, to all the charged or neutral interacting debris produced by the collision. Since collisions are observed in the centre of mass, a significant momentum imbalance may signal the presence of one or more non-interacting particles, presumably neutrinos.

The method can be conveniently implemented with calorimeters, since their energy response can be made rather uniform for different incident particles. Calorimetry is also ideally suited to the accurate measurement of the energy of the accompanying high-energy electron for process (3). Energy depositions (Fig. 7) in individual cells, Ei, are converted into an energy flow vector $\vec{E}_i=\vec{n}_iE_i$, where $\vec{n}_i$ is the unit vector pointing from the collision point to (the centre of) the cell. Then, for relativistic particles and for an ideal calorimeter response $\Sigma_i\vec{E}_i=0$, provided no non-interacting particle is emitted. The sum covers the whole solid angle. In reality there are finite residues to the sum: $\vec{\Delta E}_M=\Sigma_i\vec{E}_i$. This quantity is called the 'missing energy' vector. Obviously in the case of a neutrino emission, $\vec{p}_\nu=-\vec{\Delta E}_M$. In the case of process (3) the effect is particularly spectacular, since in the centre of mass of the W the neutrino momentum $p^*_\nu=m_W/2$ is very large.

The practical realization of such a detector [14] is shown in Fig. 8a. After momentum analysis in a large-image drift chamber in a horizontal magnetic field of 7000 G oriented normal to the beam directions, six concentric sets of finely segmented calorimeters (Fig. 8 b) surround the collision point, down to

CONSTRUCTION OF ENERGY VECTORS



$\vec{u}_i$ = unit vector

$\vec{E}_i = E_i \, \vec{u}_i$

Momentum conservation (ideal detector) $\longrightarrow \sum_i \vec{E}_i = 0$

$\Delta \vec{E}_m = \Sigma \vec{E}_i$

$\Sigma \left| E_T \right|$ = event "temperature"

Fig. 7. Principal diagram for constructing energy vectors and the missing energy of the event

angles of 0.2° with respect to the beam directions. The operation of these calorimeters is shown schematically in Fig. 9a. The first four segments are sandwiches of lead and scintillator, in which electrons are rapidly absorbed (Fig. 9b), followed by two sections of iron/scintillator sandwich (which is also. the return yoke of the magnetic field). All hadrons are completely absorbed within these calorimeters. Muons are detected by eight planes of large drift chambers which enclose the whole detector volume. If one or more muons are detected, their momenta, measured by magnetic curvature, must be added 'by hand' to the energy flow vector.

The performance of the energy flow measurement has been tested with

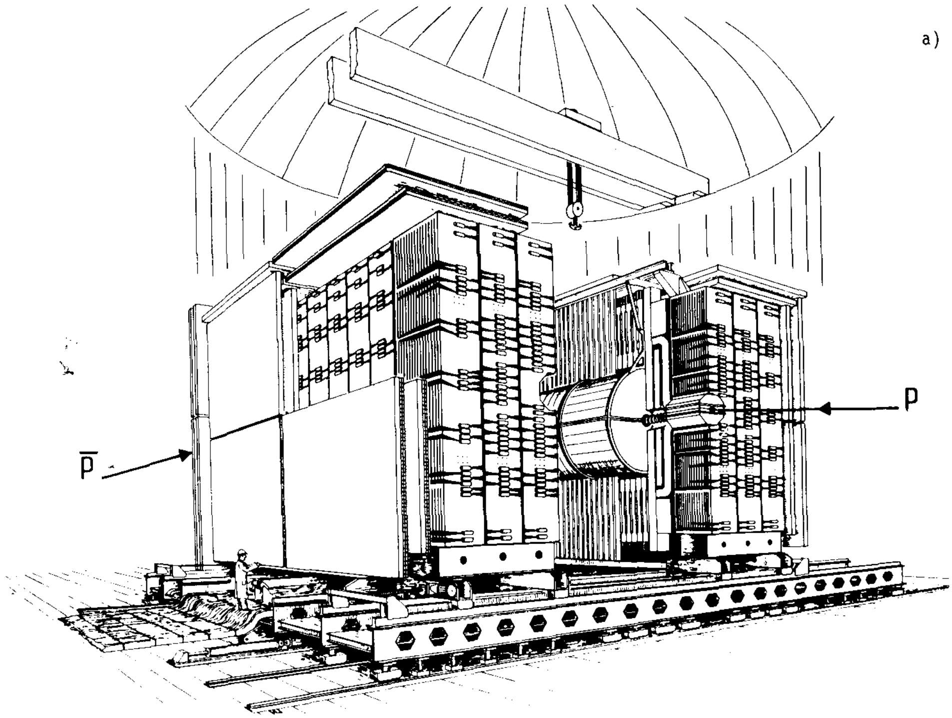


Fig. 8a. The UA1 detector solid-angle is fully covered down to 0.2°.

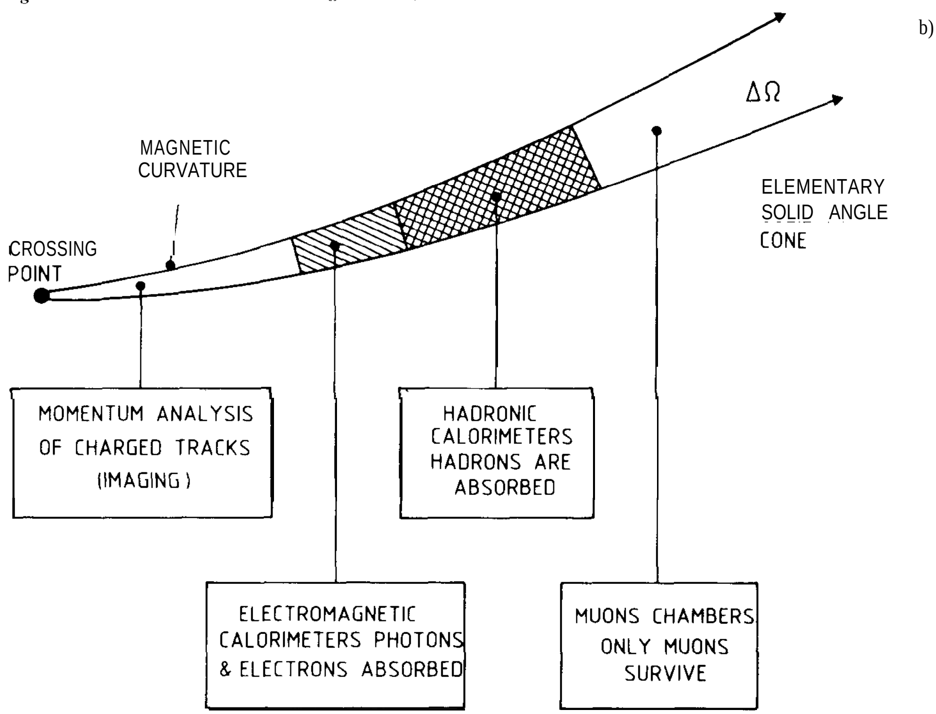


Fig. 8b. The schematic functions of each of the elementary solid-angle elements constituting the detector structure.

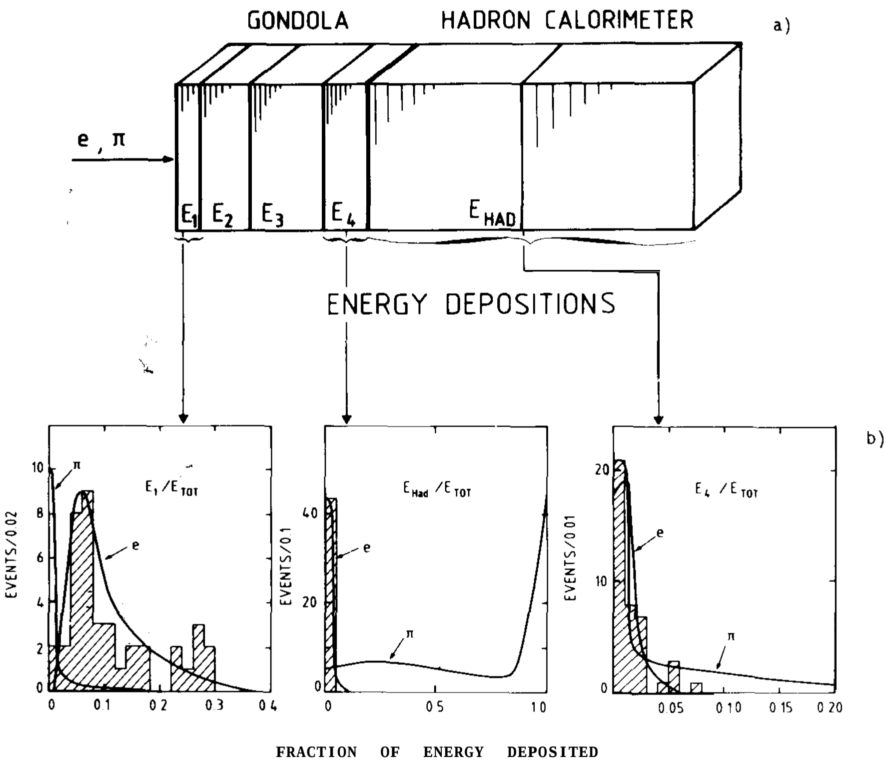


Fig. 9. a) Schema of an elementary solid-angle cell. After four segments of lead/scintillator sandwich, there are two elements of iron/scintillator sandwich, which is also the magnetic field return loop. b) Energy depositions for high-energy pions and electrons. The nature of the particle can be discriminated looking at the transition curve.

standard collisions (minimum bias). Fig. 10 shows how well the vertical component of the missing-energy vector is observed for minimum bias events. The missing energy $\overrightarrow{\Delta E}_M$ resolution for each transverse component can be parametrized as $\sigma=0.43\sqrt{\Sigma_i E_T^{(i)}}$, where $\Sigma_i E_T^{(i)}$, in units of GeV, is the scalar sum of the transverse components of the energy flow $E_T^{(i)}$. The same parametrization also holds for events which contain high transverse momentum jets, and for which the detector non-uniformities are more critical since energy deposition is highly localized (Fig. 11). The resolution function is shown in Fig. 12, where the missing energy for two-jet events is shown along with a Monte Carlo calculation of the expected distribution based on the expected behaviour of the calorimeters as determined by test-beam data and the measured fragmentation functions of jets.

For a typical event with $\Sigma_i E_T^{(i)}=80$ GeV, we measure the transverse components of $\overrightarrow{\Delta E}_M$ to about 4 GeV. The longitudinal component of the momentum balance will not be used in the present analysis since, in spite of the smallness of the window through which the beam pipes pass ($\leqslant 0.2°$), energetic particles quite often escape through the aperture.

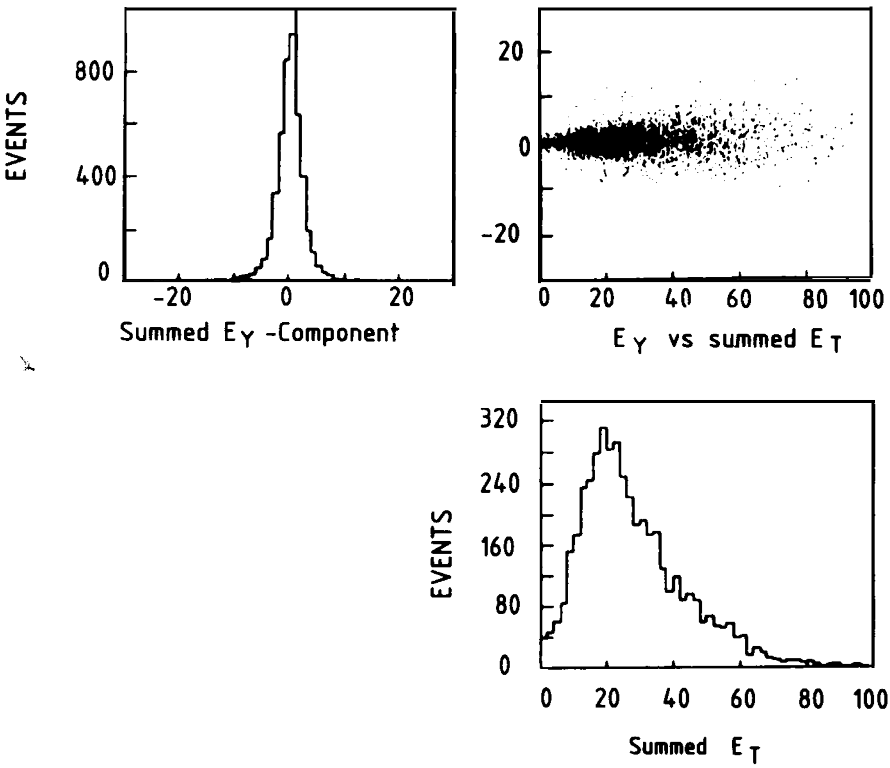


Fig. *10.* Scatter-plot of the vertical component of missing transverse energy versus the total transverse energy observed in all calorimeter cells.

5. Observation of the $W \to e + \nu$ signal

The observation by the UAl Collaboration [15] of the charged intermediate vector boson was reported in a paper published in February 1983, followed shortly by a parallel paper from the UA2 Collaboration [16]. Mass values were given: $m_W=(80\pm5)$ GeV/c^2(UA1) and $m_W=(80^{+10}_{-6})$ GeV/c^2(UA2). Since then, the experimental samples have been considerably increased, and one can now proceed much further in understanding the phenomenon. In particular, the assignment of the events to reaction (3) can now be proved rather than postulated. We shall follow here the analysis of the UAl events [17].

Our results are based on an integrated luminosity of 0.136 pb^{-1}. We first performed an inclusive search for high-energy isolated electrons. The trigger selection required the presence of an energy deposition cluster in the electromagnetic calorimeters at angles larger than 5°, with transverse energy in excess of 10 GeV. In the event reconstruction this threshold was increased to 15 GeV, leading to about 1.5×10^6 beam-beam collision events.

By requiring the presence of an associated, isolated track with $p_T>7$ GeV/c in the central detector, we reduced the sample by a factor of about 100. Next, a maximum energy deposition (leakage) of 600 MeV was allowed in the hadron calorimeter cells after the electromagnetic counters, leading to a sample of 346

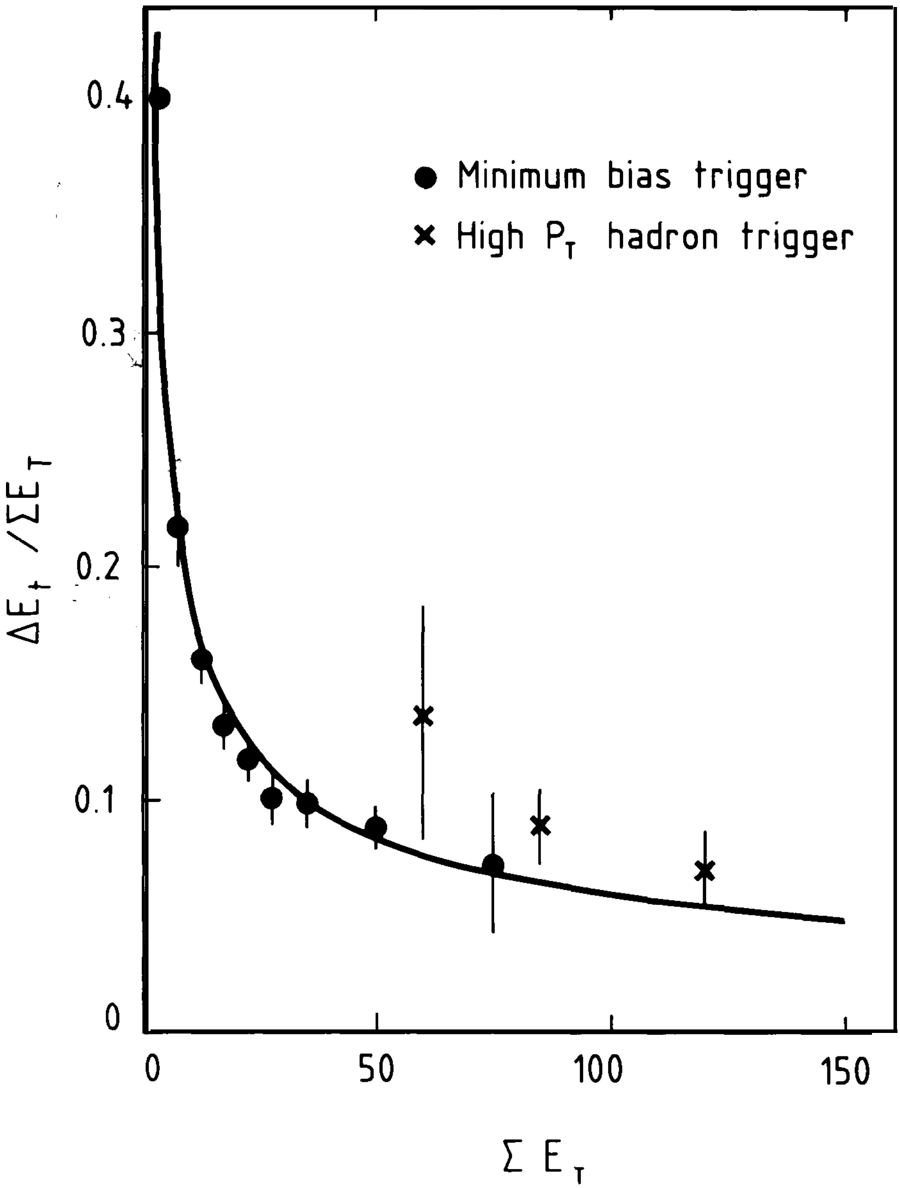


Fig. *11*. Missing-energy resolution for minimum-bias and jet events

events. We then classified events according to whether there was prominent jet activity.

We found that in 291 events there was a clearly visible jet within an azimuthal angle cone $|\Delta\phi|<30°$ opposite to the 'electron' track. These events were strongly contaminated by jet-jet events in which one jet faked the electron signature and had to be rejected. We were left with 55 events without any jet, or with a jet not back-to-back with the 'electron' within 30". These events had a very clean electron signature (Fig. 13) and a perfect matching between the point of electron incidence and the centroid in the shower detec-

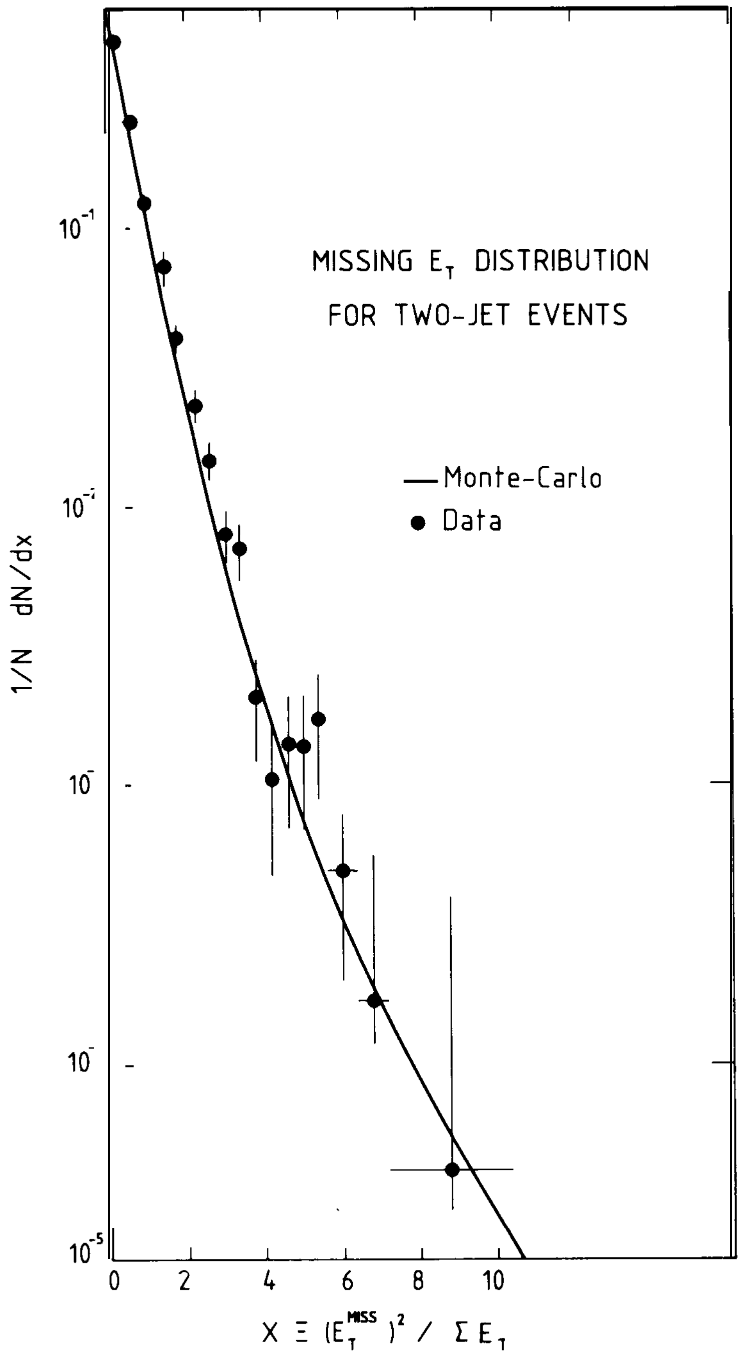


Fig. 12. Transverse energy balance observed for a sample of two-jet events. To convert the horizontal scale to the number of standard deviations (n), use the relationship $n^2 \approx 2x$. Variables have been chosen in such a way as to transform a Gaussian basic response of the calorimeters into a linear plot. The continuous line is the result of a calculation based on the expected calorimeter responses, as measured with test-beam particles.

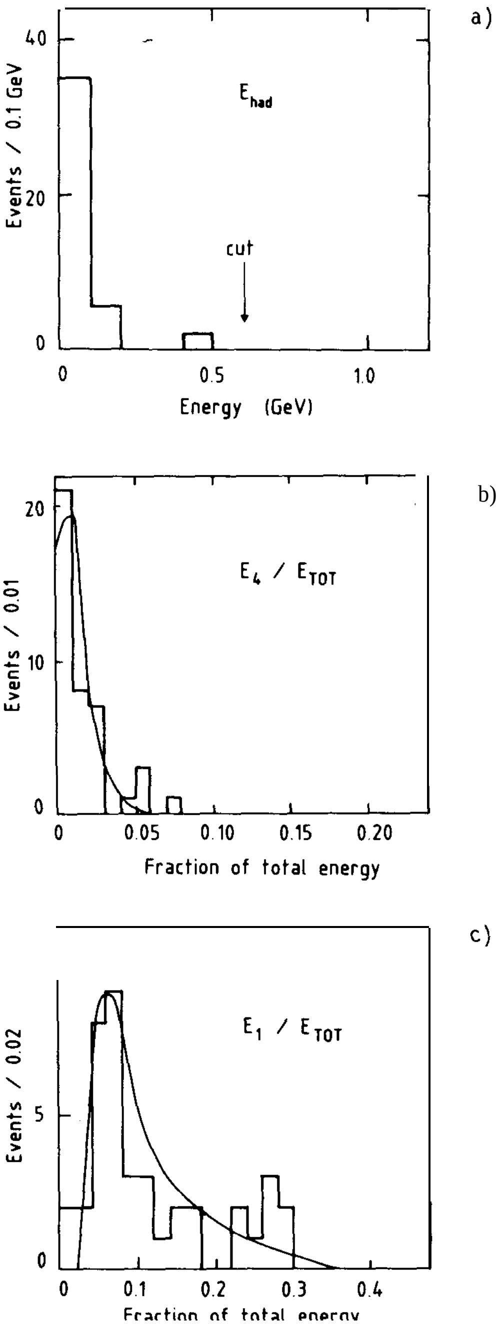


Fig. *13*. Distributions showing the quality of the electron signature:
a) The energy deposition in the hadron calorimeter cells behind the 27 radiation lengths (r. l.) of the e.m. shower detector.
b) The fraction of the electron energy deposited in the fourth sampling (6 r.l. deep, after 18 r.l. converter) of the e.m. shower detector. The curve is the expected distribution from test-beam data.
c) As distribution (b) but for the first sampling of the e.m. shower detector (first 6 r.l.).

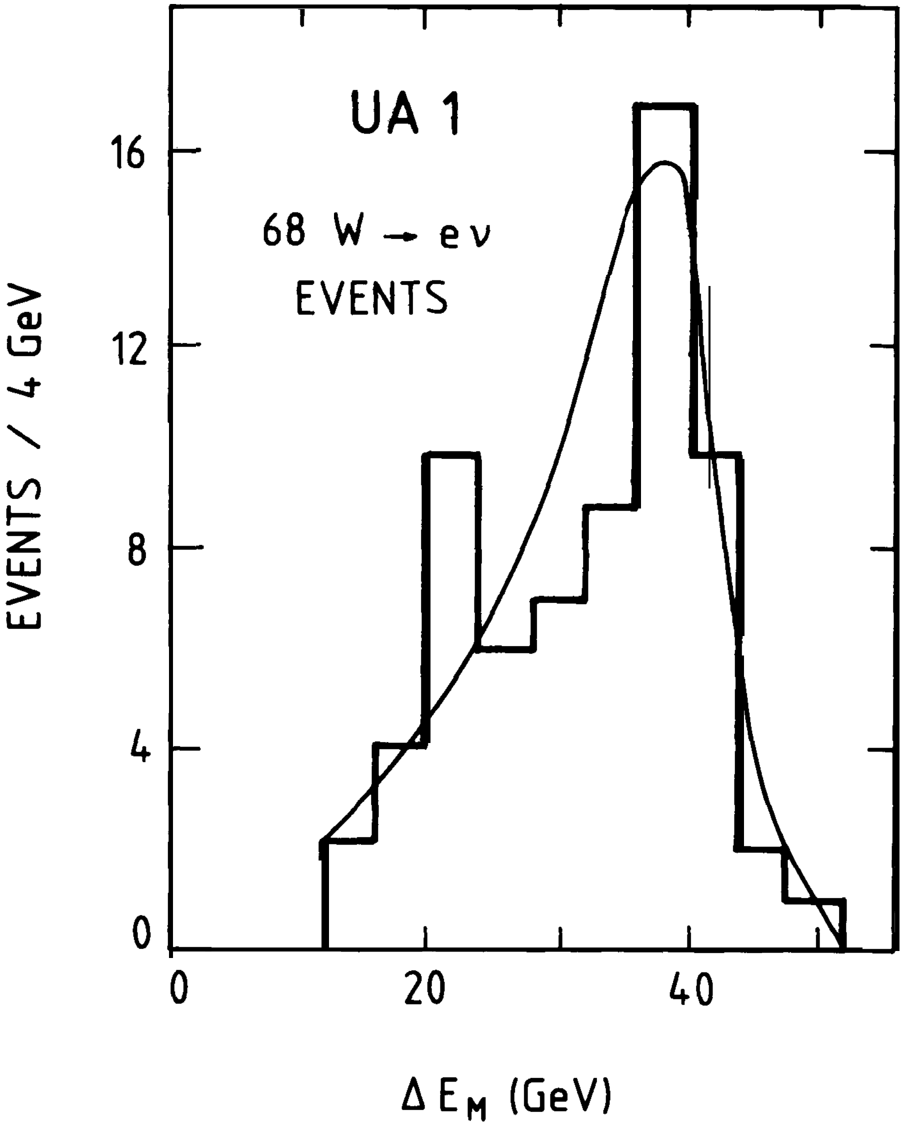


Fig. 14. The distribution of the missing transverse energy for those events in which there is a single electron with E_T>15 GeV, and no coplanar jet activity. The curve represents the resolution function for no missing energy normalized to the three lowest missing-energy events.

tors, further supporting the absence of composite overlaps of a charged track and neutral π^0's expected from jets.

The bulk of these events was characterized by the presence of neutrino emission, signalled by a significant missing energy (see Fig. 14). According to the experimental energy resolutions, at most the three lowest missing-energy events were compatible with no neutrino emission. They were excluded by the cut E_T^{miss}>15 GeV. We were then left with 52 events.

In order to ensure the best accuracy in the electron energy determination, only those events were retained in which the electron track hit the electromag-

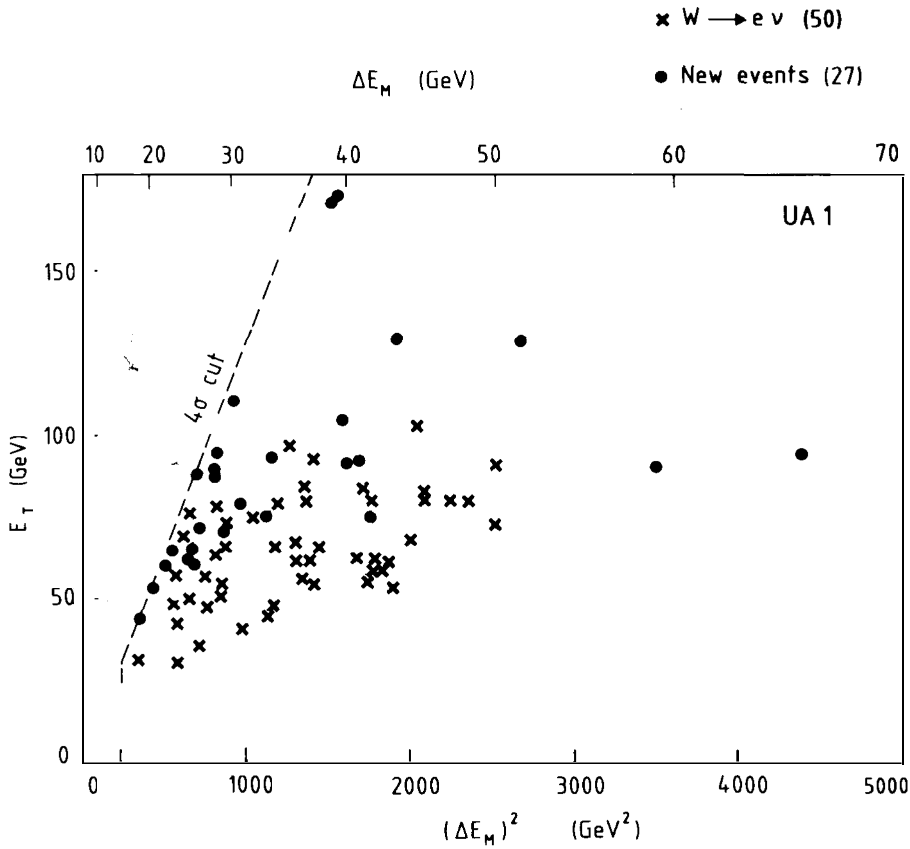


Fig. 15a. Missing transverse energy squared versus ΣE_T for all verified events which have ΔE_m more than 4 St. dev. from zero for all events with $W \to e+v$ decays removed. The events are labelled according to their topology.

netic detectors more than ±15” away from their top and bottom edges. The sample was then reduced to 43 events.

An alternative selection was carried out, based on the inclusive presence of a significant missing energy [18]. This is illustrated in Fig. 15a, where all events with missing energy in excess of 4 standard deviations are shown. One can see that previously selected electron events are found as a subset of the sample. However, a significant number of additional events (twenty-seven) were also recorded, in which there was either a jet or an electromagnetic cluster instead of the isolated electrons (Fig. 15 b). Evidently the inclusive missing-energy definition implies a broader class of physical phenomena (Fig. 16 c) than the simple $W \to e+v$ decay (Figs. 16a, b). As the study of these events [19] is beyond the scope of this lecture, it will not be pursued any further.

We proceeded to a detailed investigation of the events in order to elucidate their physical origin. The large missing energy observed in all of them was interpreted as being due to the emission of one or several non-interacting neutrinos. A very strong correlation in angle and energy was observed (in the plane normal to the colliding beams, where it could be determined accurately),

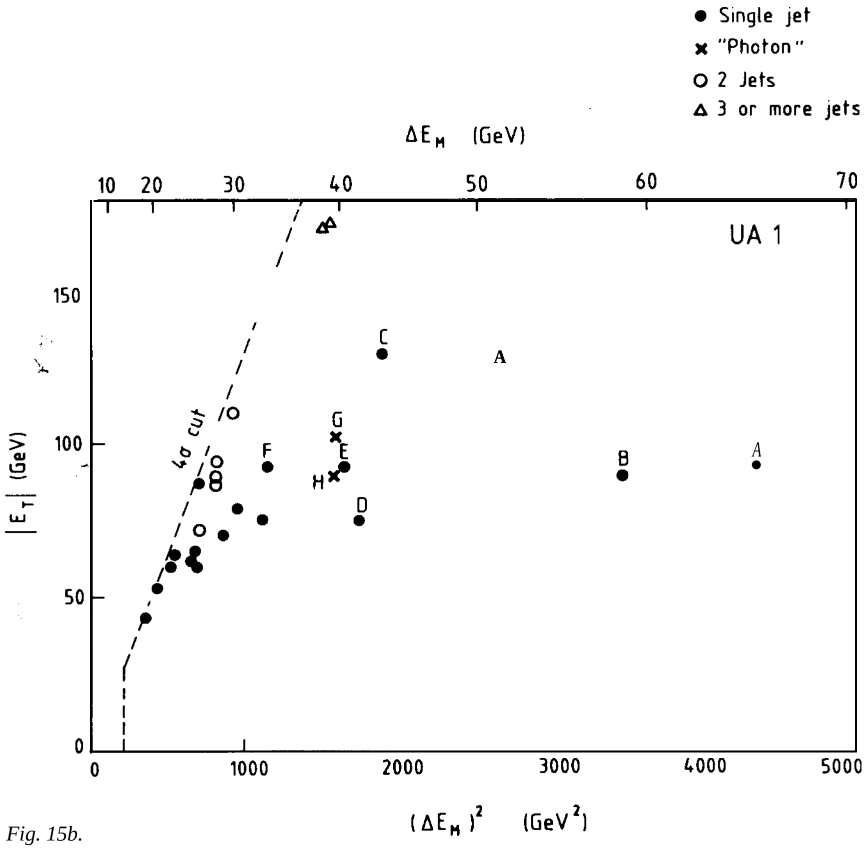


Fig. 15b.

with the corresponding electron quantities, in a characteristic back-to-back configuration expected from the decay of a massive, slow particle (Figs. 17a, b) . This suggested a common physical origin for the electron and for one or several neutrinos.

In order to have a better understanding of the transverse motion of the electron-neutrino(s) system, we studied the experimental distribution of the resultant transverse momentum $p_T^{(W)}$ obtained by adding the neutrino(s) and electron momenta (Fig. 18). The average value was $p_T^{(W)}=6.3$ GeV/c. Five events which had a visible jet had also the highest values of $p_T^{(W)}$. Transverse momentum balance was almost exactly restored when the vector momentum of the jet was added. The experimental distribution was in good agreement with the many theoretical expectations from quantum chromodynamics (QCD) for the production of a massive state via the Drell-Yan quark-antiquark annihilation [20]. The small fraction (10 %) of events with a jet were then explained as hard gluon bremsstrahlung in the initial state.

Several different hypotheses on the physical origin of the events were tested by looking at kinematical quantities constructed from the transverse variables of the electron and the neutrino(s). We retained two possibilities, namely: i) the two-body decay of a massive particle into the electron and one neutrino,

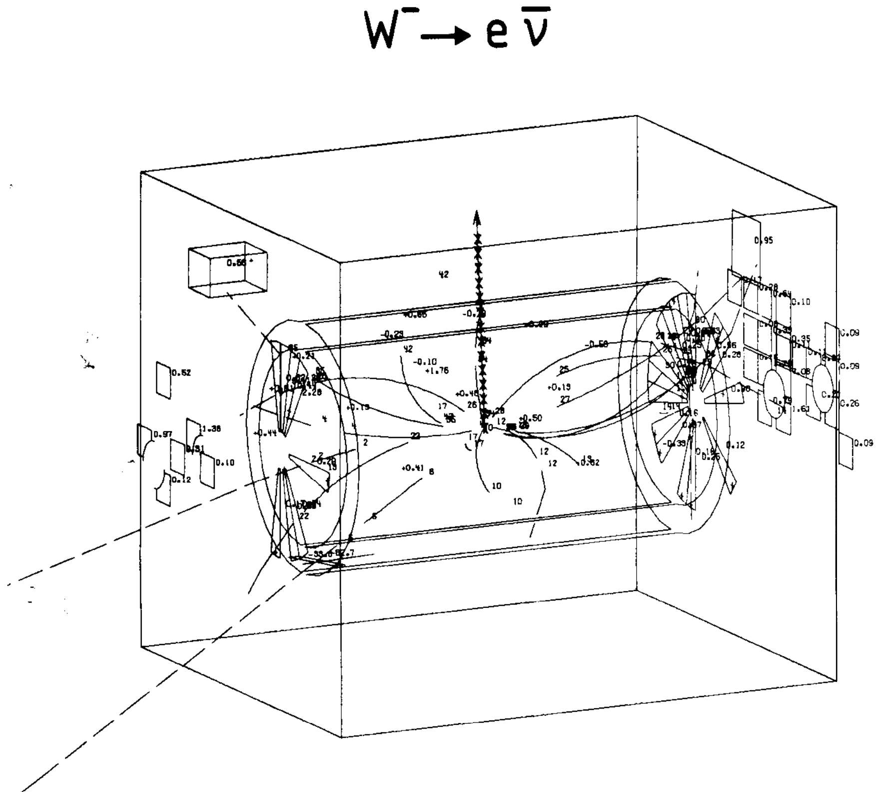


Fig. 16a. Event of the type $W^- \rightarrow e^- + \bar{\nu}_e$. All tracks and calorimeter cells are displayed.

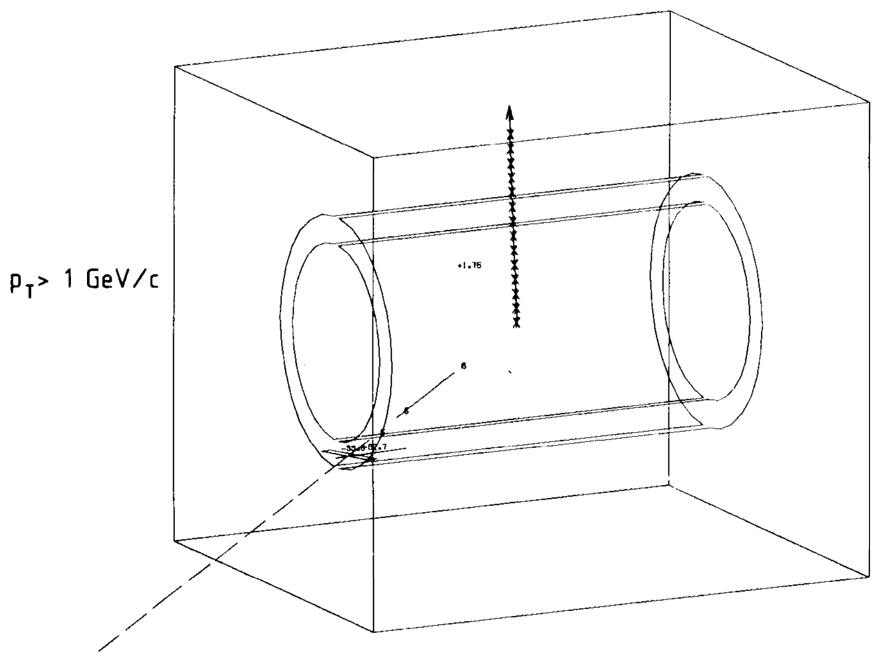


Fig. *16b*. The same as picture (a), except that now only particles with $p_T>1$ GeV/c and calorimeters with $E_T>1$ GeV are shown.

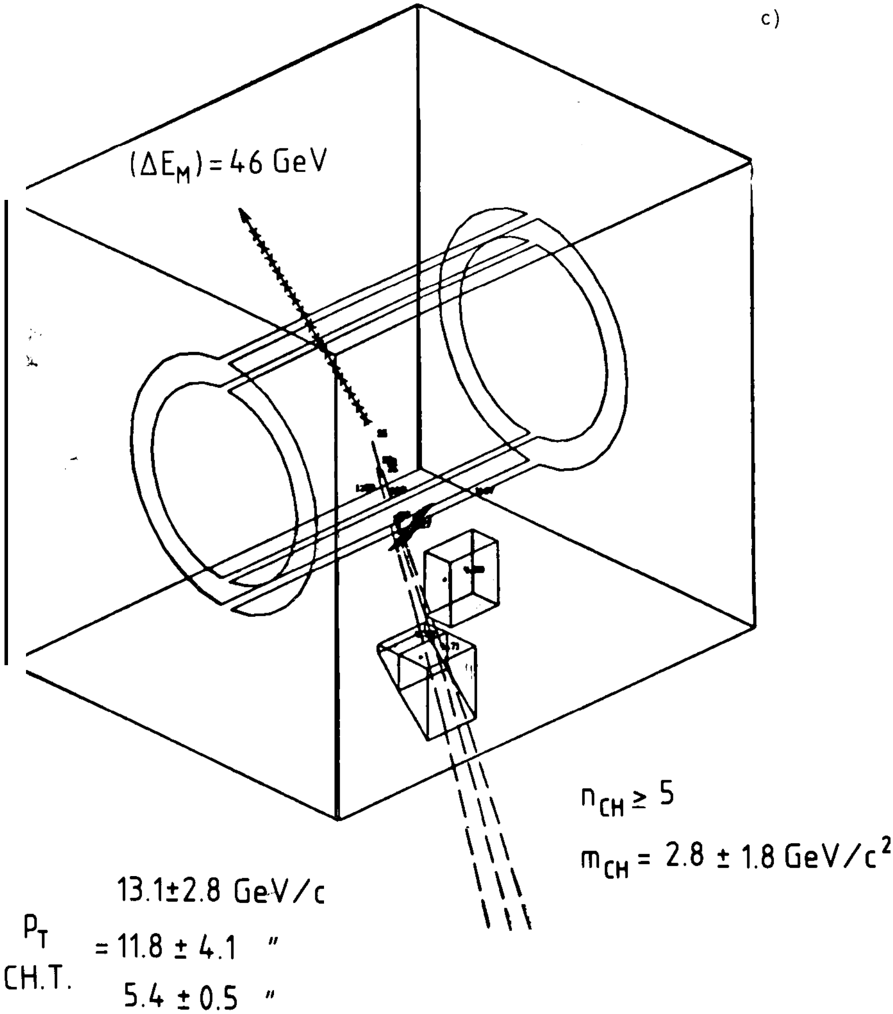


Fig. 16c. Event of the type, jet+missing energy. Only tracks with pT>1.5 GeV/c and cells with E_T> 1.0 GeV are displayed.

$W \rightarrow e+\nu_e$; and ii) the three-body decay into two, or possibly more, neutrinos and the electron. It can be seen from Figs. 19 a and 19 b that hypothesis (i) is strongly favoured. At this stage, the experiment could not distinguish between one or several closely spaced massive states.

With the help of a sample of isolated hadrons at large transverse momenta, we estimated in detail the possible sources of background coming from ordinary hadronic interactions, and we concluded that they were negligible (CO.5 events). (For more details on background, we refer the reader to Ref. 20.) However, we expect to get some background events from other decays of the W, namely:

$$W \rightarrow \tau+\nu_\tau[\tau \rightarrow \pi^{\pm}(\pi^0)+\nu_\tau] \qquad (<0.5 \text{ events})$$

or

$$W \rightarrow \tau+\nu_\tau(\tau \rightarrow e+\nu_e+\nu_\tau) \qquad (= 2 \text{ events}).$$

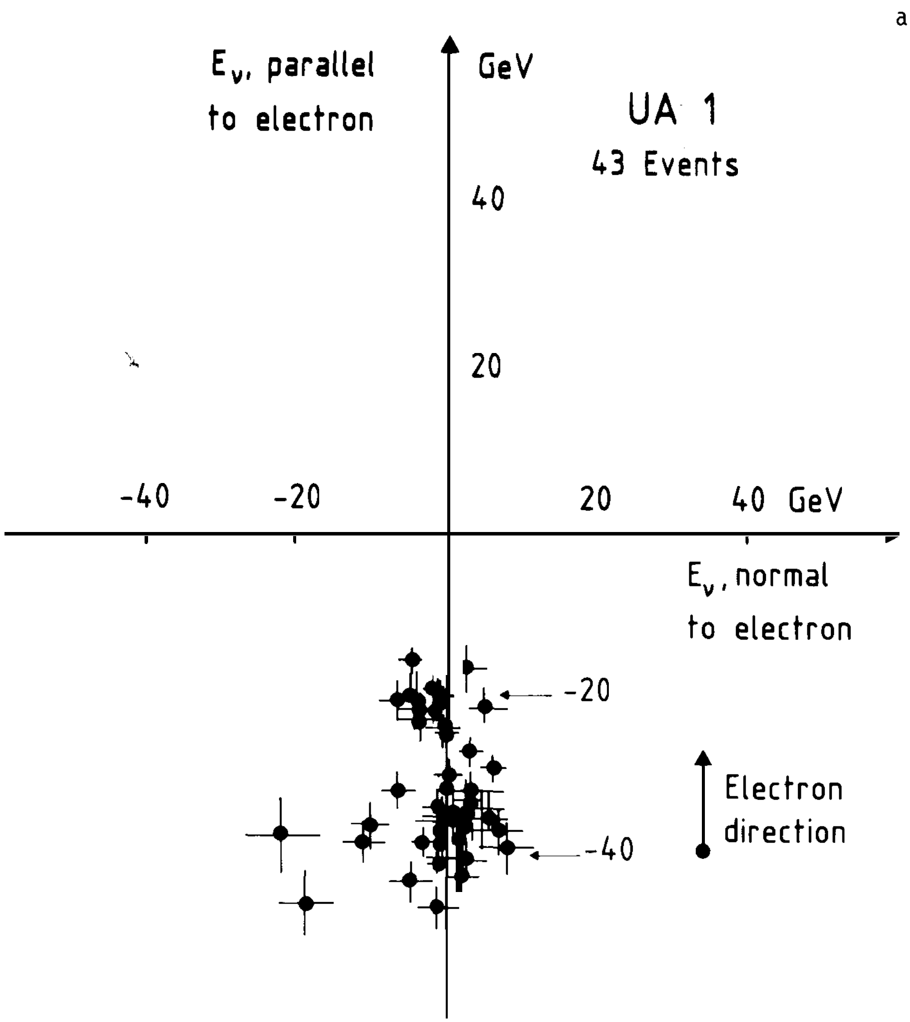


Fig. 17a. Two-dimensional plot of the transverse components of the missing energy (neutrino momentum). Events have been rotated to bring the electron direction to point along the vertical axis. The striking back-to-back configuration of the electron-neutrino system is apparent.

These events were expected to contribute at only the low-p_T part of the electron spectrum, and could even be eliminated in a more restrictive sample.

A value of the W mass can be extracted from the data in a number of ways:

i) It can be obtained from the inclusive transverse momentum distribution of the electrons (Fig. 19 a), but the drawback of this technique is that the transverse momentum of the W particle must be known. Taking the QCD predictions [21], in reasonable agreement with experiment, we obtained $m_W=(80.5\pm0.5)$ GeV/c^2.

ii) We can define a transverse mass variable, $m_T^2=2p_T^{(e)}p_T^{(\nu)}(1-\cos\phi)$, with the property smw, where the equality would hold for only those events with no longitudinal momentum components. Fitting Fig. 19b to a

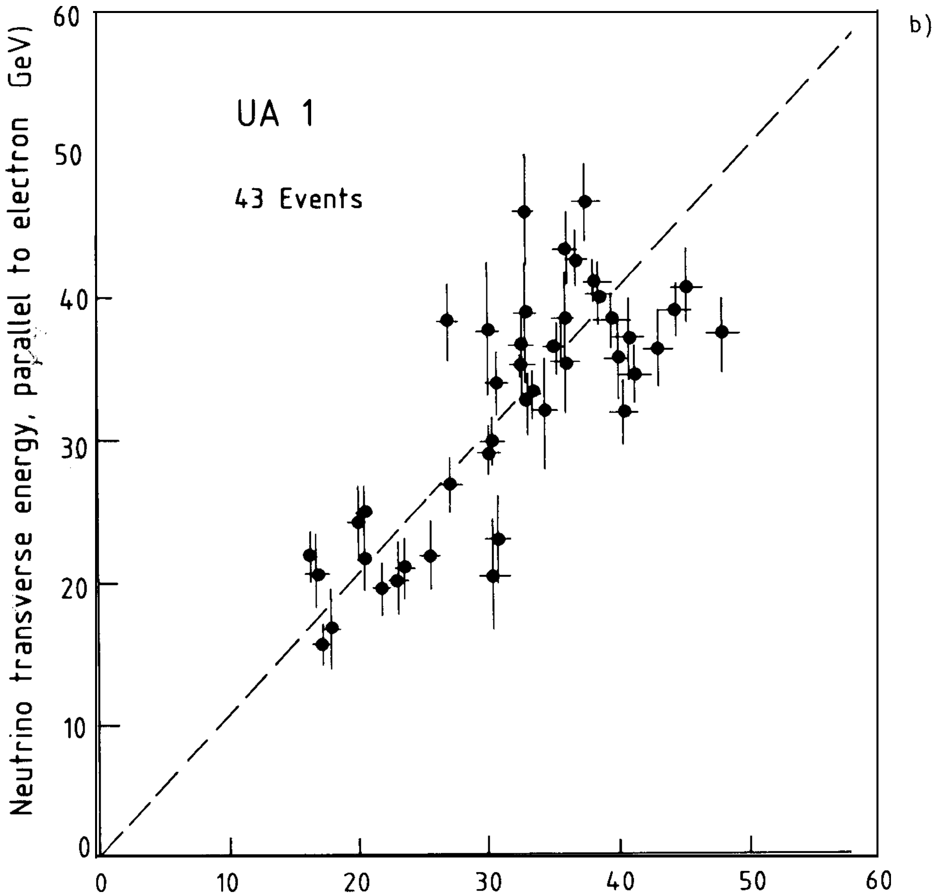


Fig. *17b*. Correlation between the electron and neutrino transverse energies. The neutrino component along the electron direction is plotted against the electron transverse **energy.**

common value of the mass was done almost independently of the transverse motion of the W particles, $m_w = (80.3^{+0.4}_{-1.3})$ GeV/c^2. It should be noted that the lower part of the distribution in $m_T^{(W)}$ was slightly affected by $W \rightarrow \tau + \nu_\tau$ decays and other backgrounds.

iii) We can define an enhanced transverse mass distribution, selecting only events in which the decay kinematics is largely dominated by the transverse variable with the simple cuts $p_T^{(e)}, p_T^{(\nu)} > 30$ GeV/c. The resultant distribution (Fig. 19c) then showed a relatively narrow peak at approximately 76 GeV/c^2. Model-dependent corrections now only contributed to the difference between this average mass value and the fitted m_w value, m_w = (80.9±1.5) GeV/c^2. An interesting upper limit to the width of the W was also derived from the distribution, namely $\Gamma_T \leqslant 7$ GeV/c^2(90% confidence level).

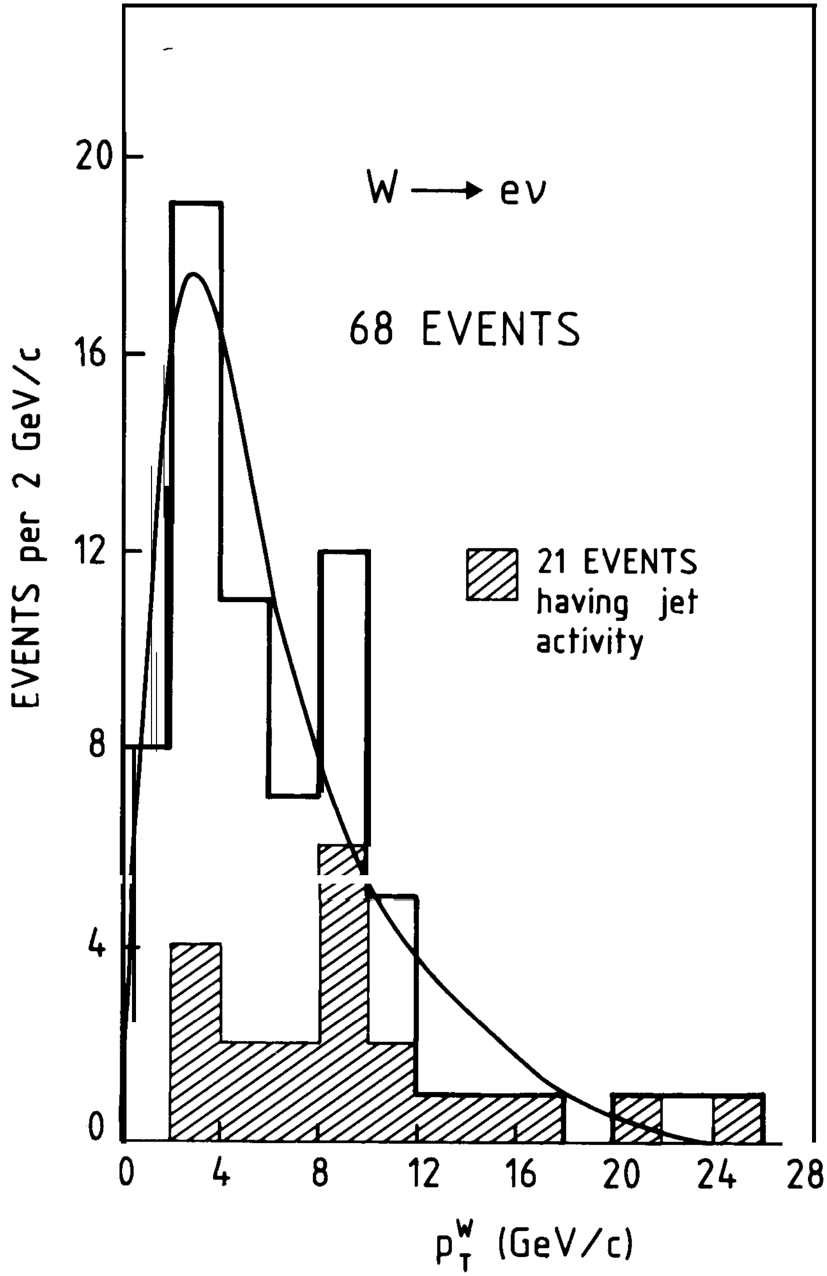


Fig. 18. The transverse momentum distribution of the W derived from our events using the electron and missing transverse energy vectors. The highest $p_T^{(W)}$ events have a visible jet (shown in black in the figure). The data are compared with the theoretical predictions for W production based on QCD (Ref. [21]).

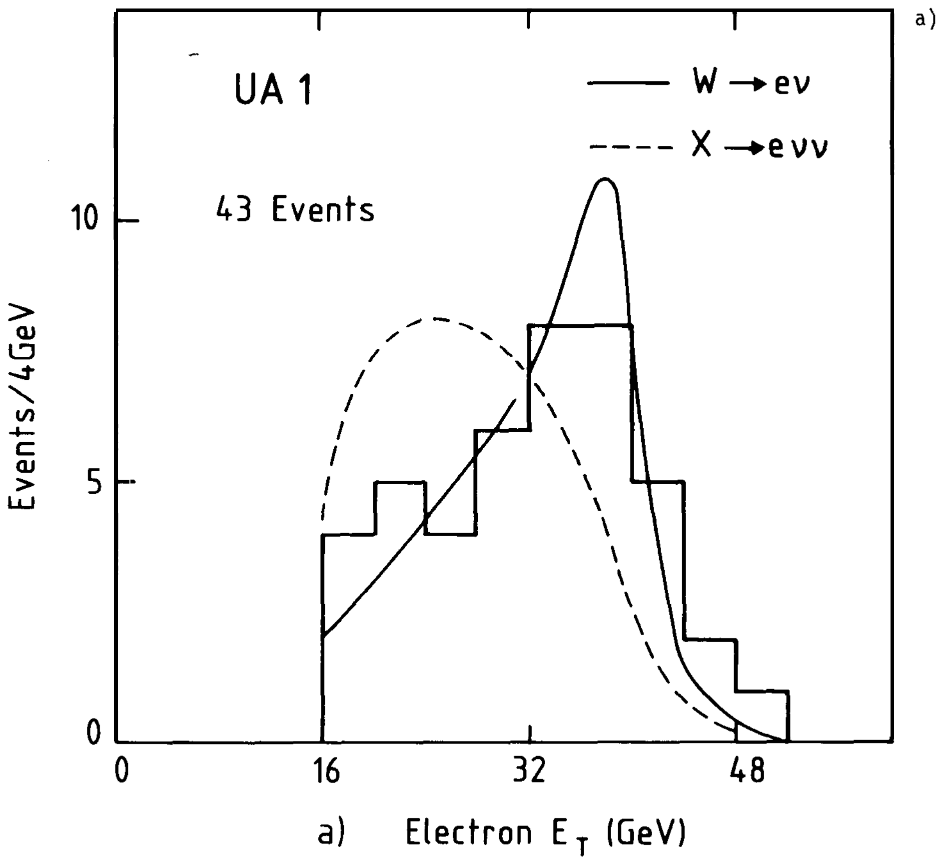


Fig. *19a.* The electron transverse energy distribution. The two curves show the results of a lit of the enhanced transverse mass distribution to the hypotheses $W \to e+\nu$ and $X \to e+\nu+\nu$. The first hypothesis is clearly preferred.

The three mass determinations gave very similar results. We preferred to retain the result of method (iii), since we believed it to be the least affected by systematic effects, even if it gave the largest statistical error. Two important contributions had to be added to the statistical errors:

i) *Counter-to-counter calibrations.* They were estimated to be 4% r.m.s. In the determination of the W mass this effect was greatly attenuated to a negligible level, since many different elements contributed to the event sample.

ii) *Calibration of the absolute energy scale.* This was estimated to be ± 3 %, and of course affects both the Z^0 and the W samples by the same multiplicative factor.

Once the decay reaction $W \to e+\nu_e$ was established, the longitudinal momentum of the electron-neutrino system was determined with a twofold ambiguity for the unmeasured longitudinal component of the neutrino momentum. The overall information of the event was used to establish momentum and energy conservation bounds in order to resolve this ambiguity in 70% of the cases. Most of the remaining events had solutions which were quite close,

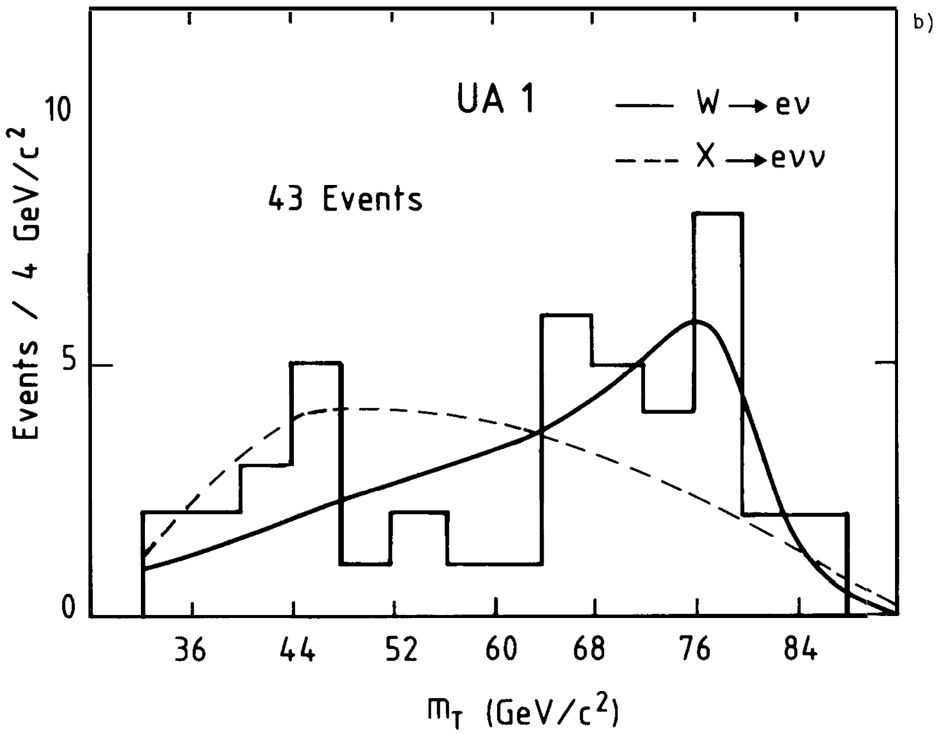


Fig. **19b.** The distribution of the transverse mass derived from the measured electron and neutrino vectors. The two curves show the results of a lit to the hypotheses $W \rightarrow e+\nu$ and $X \rightarrow e+\nu+\nu$.

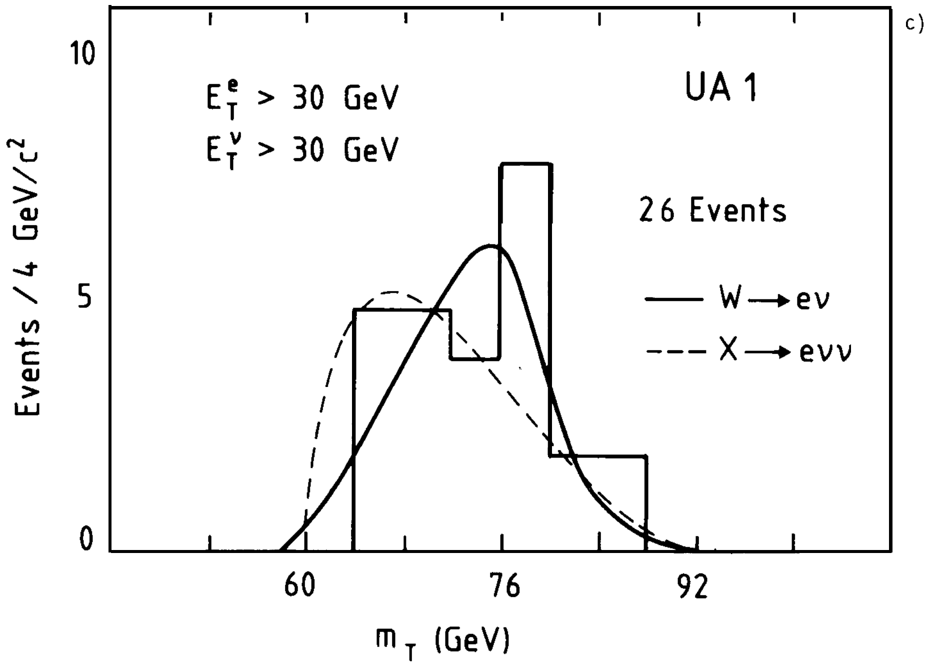


Fig. **19c.** The enhanced electron-neutrino transverse mass distribution (see text). The two curves show the results of a fit to the hypotheses $W \rightarrow e + \nu$ and $X \rightarrow e + \nu + \nu$.

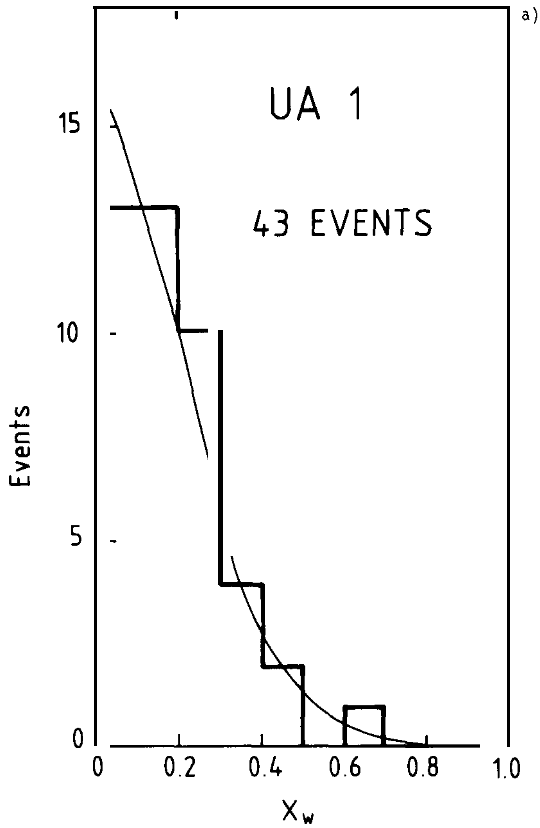


Fig. 20a. The fractional beam energy x_w carried by the W. The curve is the prediction obtained by assuming that the W has been produced by $q\bar{q}$ fusion. Note that in general there are two kinematic solutions for x_w (see text), which are resolved in 70 % of the events by consideration of the energy flow in the rest of the event. Where this ambiguity has been resolved, the preferred kinematic solution has been the one with the lowest x_w. In the 30 % of the events where the ambiguity is not resolved, the lowest x_w solution has therefore been chosen.

and the physical conclusions were nearly the same for both solutions. The fractional beam energy x_w carried by the W particle is shown in Fig. 20a, and it appears to be in excellent agreement with the hypothesis of W production in $q\bar{q}$ annihilation [22]. Using the well-known relations $x_W = x_p - x_{\bar{p}}$ and $x_p \cdot x_{\bar{p}} = m_W^2/s$, we determined the relevant parton distributions in the proton and antiproton. It can be seen that the distributions are in excellent agreement with the expected x distributions for quarks and antiquarks in the proton and antiproton, respectively (Figs. 20b and 20c). Contributions of the u and d quarks were also neatly separated by looking at the charges of produced W events, since $(u\bar{d}) \rightarrow W^+$ and $(\bar{u}d) \rightarrow W^-$ (Figs. 20d and 20e).

6. Observation of the parity (charge conjugation) violation, and determination of the spin of the W particle

One of the most relevant properties of weak interactions is the violation of parity and charge conjugation. Evidently the W particle, in order to mediate

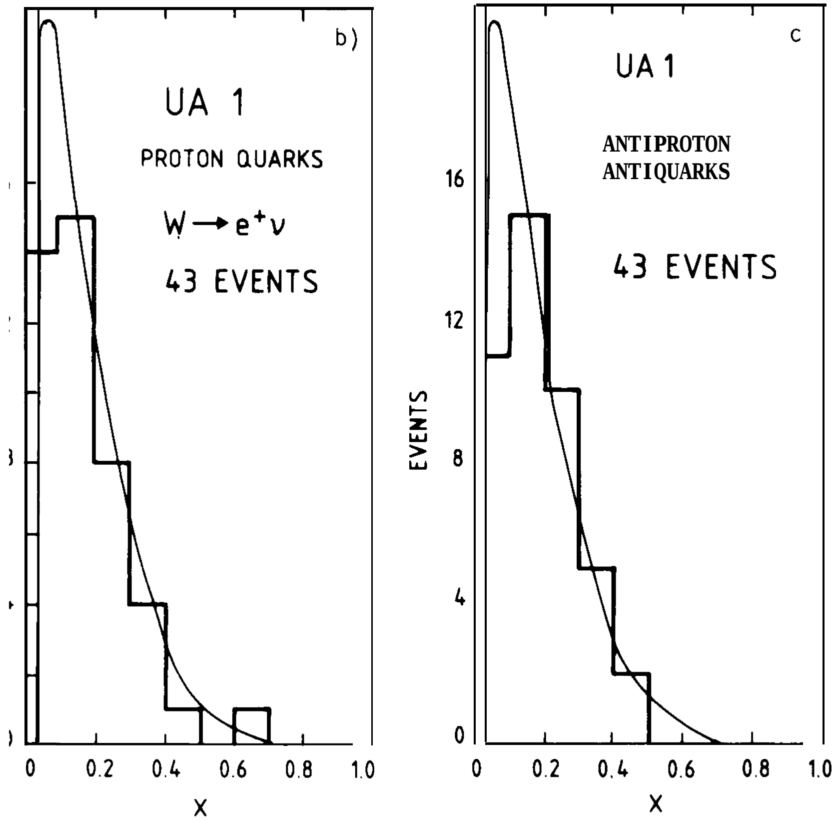
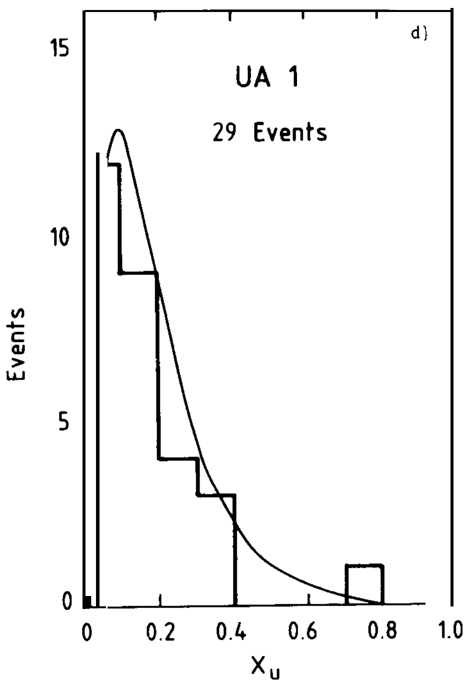


Fig. 206. The x-distribution of the proton quarks producing the W by qq fusion. The curve is the prediction assuming qq fusion.
Fig. 20c. The same as Fig. 20 b for the antiproton quarks.

weak processes, must also exhibit these properties. Furthermore, as already mentioned, the V-A nature of the four-fermion interaction implies the assignment J= 1 for its spin. Both of these properties must be verified experimentally. According to the V-A theory, **weak** interactions should act as a longitudinal polarizer of the W particles, since quarks (antiquarks) are provided by the proton (antiproton) beam. Likewise, decay angular distributions from a polarizer are expected to have a large asymmetry, which acts as a polarization analyser. A strong backward-forward asymmetry is therefore expected, in which electrons (positrons) prefer to be emitted in the direction of the proton (antiproton). In order to study this effect independently of W-production mechanisms, we have looked at the angular distribution of the emission angle θ^* of the electron (positron) with respect to the proton (antiproton) direction in the W centre of mass. Only events with no reconstruction ambiguity can be used. We verified that this does not bias the distribution in the variable cos θ^*. According to the expectations of V-A theory the distribution should be of the type $(1+\cos\theta^*)^2$, in excellent agreement with the experimental data (Fig. 21). The parity violation parameters and the spin of the W particle can be



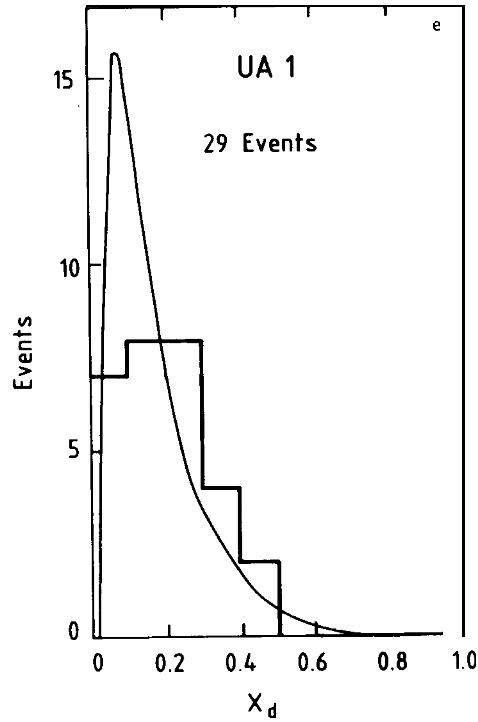


Fig. 20d. The same as Fig. 20 b but for $u(\bar{u})$ quarks in the proton (antiproton).
Fig. 20e. The same as Fig. 20 b but for d(ii) quarks in the proton (antiproton).

determined directly. It has been shown by Jacob [23] that for a particle of arbitrary spin J, one expects

$$\langle \cos \theta^* \rangle = \langle \lambda \rangle \langle \mu \rangle / J(J+1),$$

where (μ) and $\langle \lambda \rangle$ are the global helicity of the production system (ud) and of the decay system (ev), respectively.

For V-A, we then have $\langle \lambda \rangle = \langle \mu \rangle = -1$, J= 1, leading to the maximal value $\langle \cos \theta^* \rangle = 0.5$. For J=O it is obvious that $(\cos \theta^* \rangle = 0$; and for any other spin value $J \geqslant 2$,$\langle \cos \theta^* \rangle \leq 1/6$. Experimentally we find $\langle \cos\theta^* \rangle = 0.5 \pm 0.1$, which supports *both* the J= 1 assignment *and* maximal helicity states at production and decay. Note that the choice of sign $\langle \mu \rangle = (h) = \pm 1$ cannot be separated, i.e. right- and left-handed currents, both at production and decay, cannot be resolved without a polarization measurement.

7. Total cross-section and limits to higher mass W's

The integrated luminosity of the experiment was 136 nb^{-1}, and it is known to about ± 15 % uncertainty. In order to get a clean $W \rightarrow e\nu_e$, sample we selected 47 events with $p_T^{(e)} > 20$ GeV/c. The $W \rightarrow \tau\nu_\tau$ contamination in the sample was estimated to be 2±2 events. The event acceptance was computed to be 0.65, primarily because of i) the $p_T^{(e)} > 20$ GeV/ccut (0.80); ii) the jet veto require-

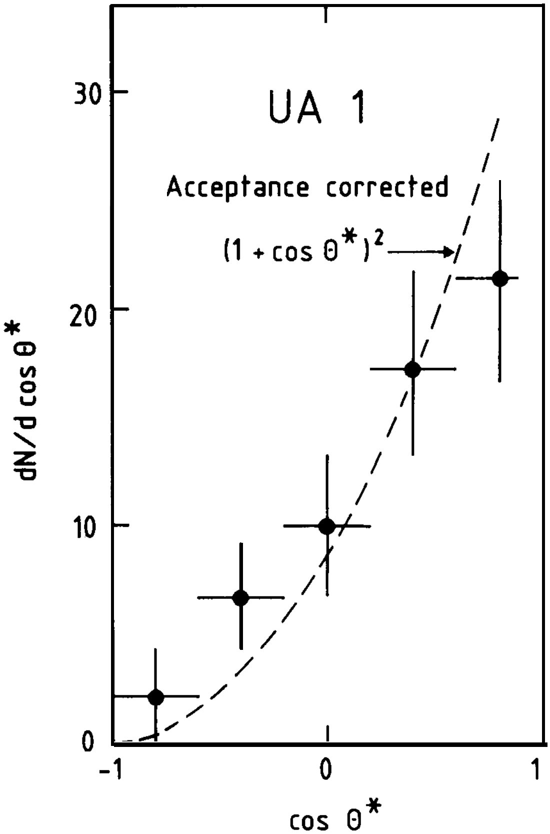


Fig. 21. The angular distribution of the electron emission angle θ^* in the rest frame of the W after correction for experimental acceptance. Only those events have been used in which the electron charge is determined and the kinematic ambiguity (see text) has been resolved. The latter requirement has been corrected for in the acceptance calculation.

ment within $\Delta\phi=\pm30°$ (0.96±0.02); iii) the electron-track isolation requirement (0.90±0.07); and iv) the acceptance of events due to geometry (0.94±0.03). The cross-section was then

$$(\sigma\cdot B)_W = 0.53\pm0.08\ (\pm0.09)\ \text{nb},$$

where the last error takes into account systematic errors. This value is in excellent agreement with the expectations for the Standard Model [22]: $(\sigma\cdot B)_W=0.39$ nb.

No event with $p_T^{(e)}$ or $p_T^{(\nu)}$ in excess of the expected distribution for $W\rightarrow e\nu$ events was observed. This result can be used to set a limit to the possible existence of very massive W-like objects (W') decaying into electron-neutrino pairs. We found $(\sigma\cdot B)_{W'}\leq30$ pb at 90% confidence level, corresponding to $m_{W'}>170$ GeV/c^2, when standard couplings and quark distributions were used to evaluate the cross-sections.

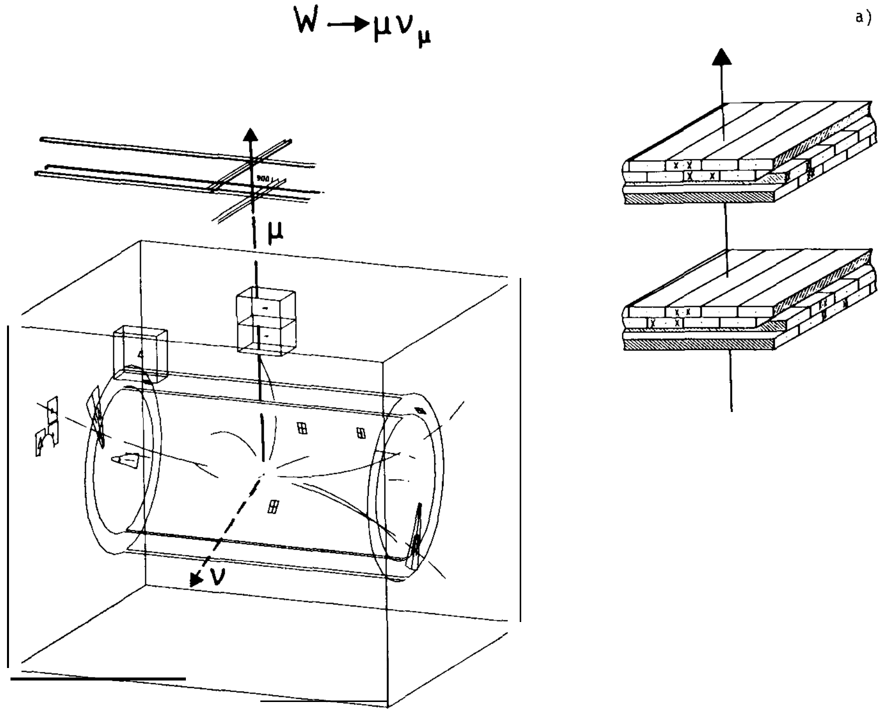


Fig. 22. Examples of decay modes of the W particle:
a) $W \rightarrow \mu+\nu_\mu$; b) $W \rightarrow \tau+\nu_\tau$; c) $W \rightarrow c+\bar{s}$; d) $W \rightarrow$ t+b $(t \rightarrow b+e+\nu)$. For the events of type (d), one can reconstruct the invariant masses of the W particle and of the decaying t-quark jet (Fig. 22 e).

8. Universality of the W coupling

The W field should exhibit a universal coupling strength for all the fundamental lepton doublets and all the quark doublets. This implies-apart from small phase-space corrections-equality of the branching ratios of the decay processes

$$W \rightarrow e\nu_e, \tag{4 a}$$

$$W \rightarrow \mu\nu_\mu, \tag{4 b}$$

$$W \rightarrow \tau\nu_\tau. \tag{4c}$$

Likewise, in the case of the quark decay channels

$$W \rightarrow ud_C, \tag{4d}$$

$$W \rightarrow cs_c, \tag{4e}$$

$$W \rightarrow tb_c, \tag{4f}$$

where t is the sixth quark (top quark) provided it exists within the kinematic range of reaction (4f). Neglecting phase-space corrections, which are probably

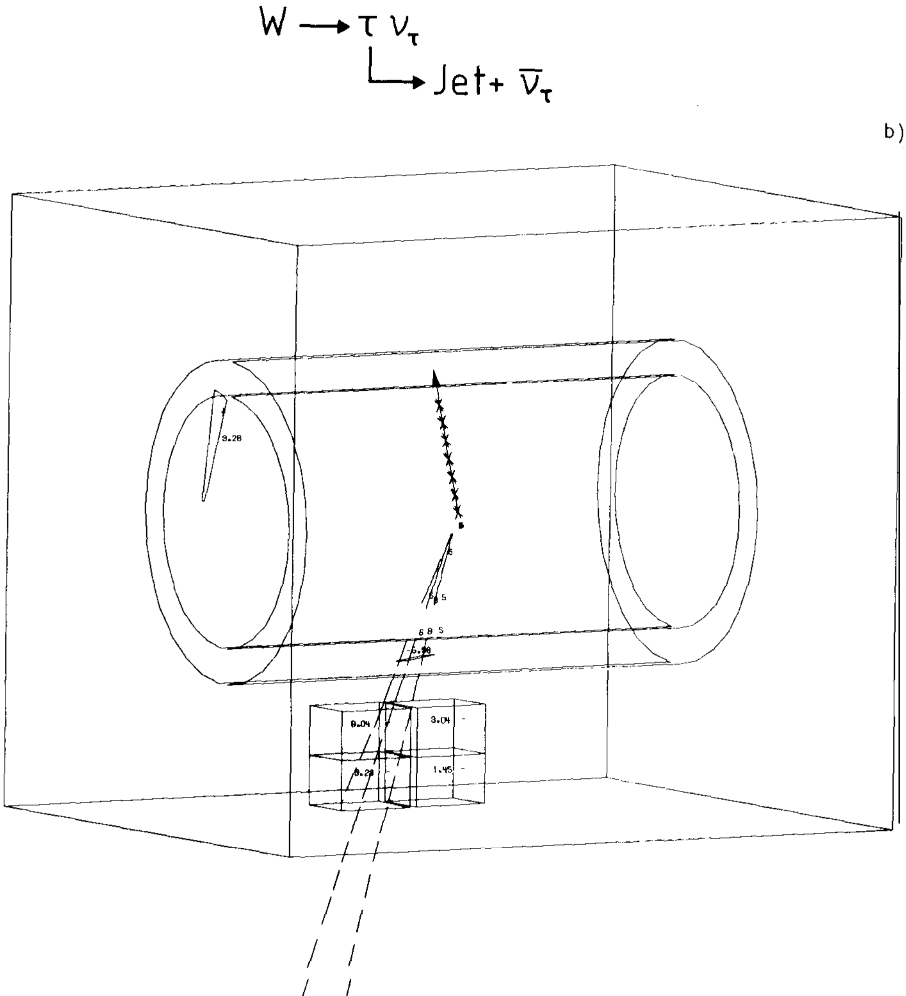


Fig. 226.

important for reaction (4f), we expect equality of the branching ratios, with an overall factor of 3 of enhancement with respect to leptonic channels [(4a) to (4c)] due to colour counting. The subscript C in channels (4d) to (4f) indicates the presence of the Cabibbo mixing. Reactions (4a) and (4d) are implied by the results of Section 5. Reactions (4 b), (4c), and (4e) have been observed, and within about ±20 % they appear to have the correct branching ratios. Some events which are believed to be evidence for the process (4f) have also been reported [24]. They are interpreted for the reaction

$$W \to t + \bar{b}_C (t \to b_C + l + \nu) \qquad (l \equiv \text{electron or muon}).$$

The $\bar{b}_C$ and b_c quarks are 'hadronized' into jets. Data are roughly consistent with $m_t \simeq 40$ GeV/c^2. Examples of reactions (4 b), (4 c), (4 e), and (4f) are shown in Figs. 22a-d, respectively.

Therefore, within the limited statistics there is evidence for universality.

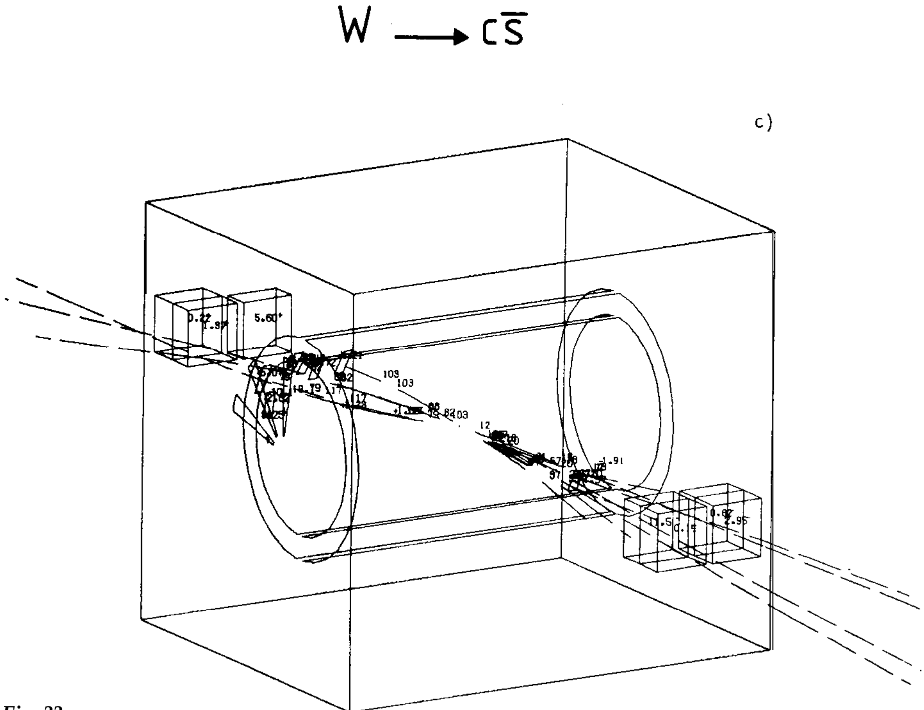


Fig. 22c.

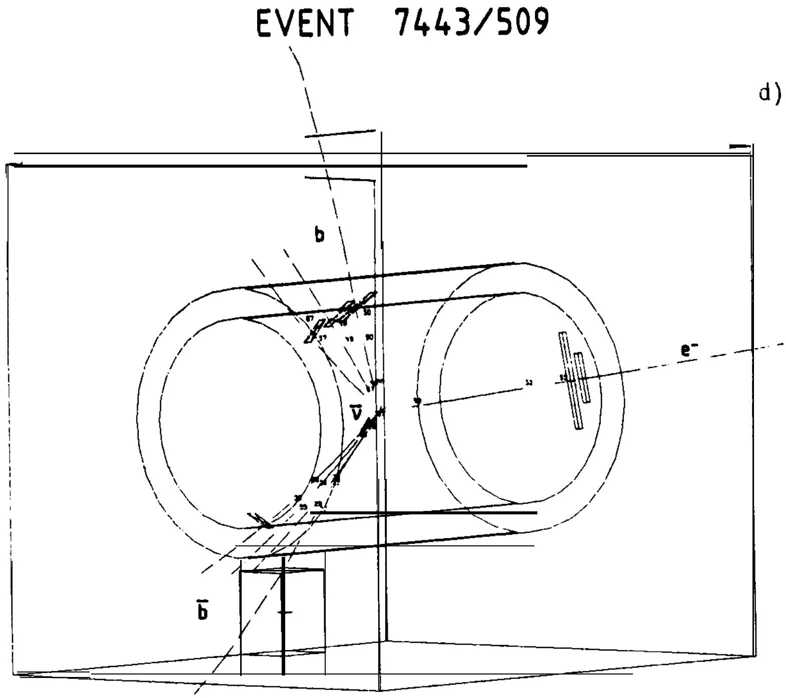
9. Can we derive weak interactions from W-particle observations?

A number of properties of weak interactions as determined by low-energy experiments can now be explained as a consequence of the experimentally observed properties of the W particles. Indeed we know that $W^{\pm}$ must couple to valence quarks at production and to (e_{ν}) pairs at decay, which implies the existence of the beta-decay processes $n \rightarrow p + e^{-} + \nu_{e}$ and $(p) \rightarrow (n) + e^{+} + \nu_{e}$. The mass value m_{w} and the cross-section measurement can then be used to calculate G_F, the Fermi coupling constant: $G_F = (1.2 \pm 0.1) \times 10^{-5}\ GeV^{-2}$. Thus the W-pole saturates the observed weak interaction rate. The interaction must be vector since J= 1, and parity is maximally violated since $\langle\mu\rangle = \langle\lambda\rangle = \pm 1$. The only missing element is the separation between V+A and V-A alternatives. For this purpose a polarization measurement is needed. It may be accomplished in the near future by studying, for instance, the decay $W \rightarrow \tau + \nu_{\tau}$ and using the τ decay as the polarization analyzer or producing intermediate vector bosons (IVBs) with longitudinally polarized protons.

The universality of couplings and the decay modes of particles of different flavours into different lepton families can also be expected on the basis of the observations of the other decay modes of the W particles.

10. *Observation of the neutral boson Z^0*

We extended our search to the neutral partner Z^0, responsible for neutral currents. As in our previous work, production of IVBs was achieved with proton-antiproton collisions at $\sqrt{s} = 540$ GeV in the UAl detector, except


EVENT 7443/509
d)
b
e⁻
ν̄
b̄

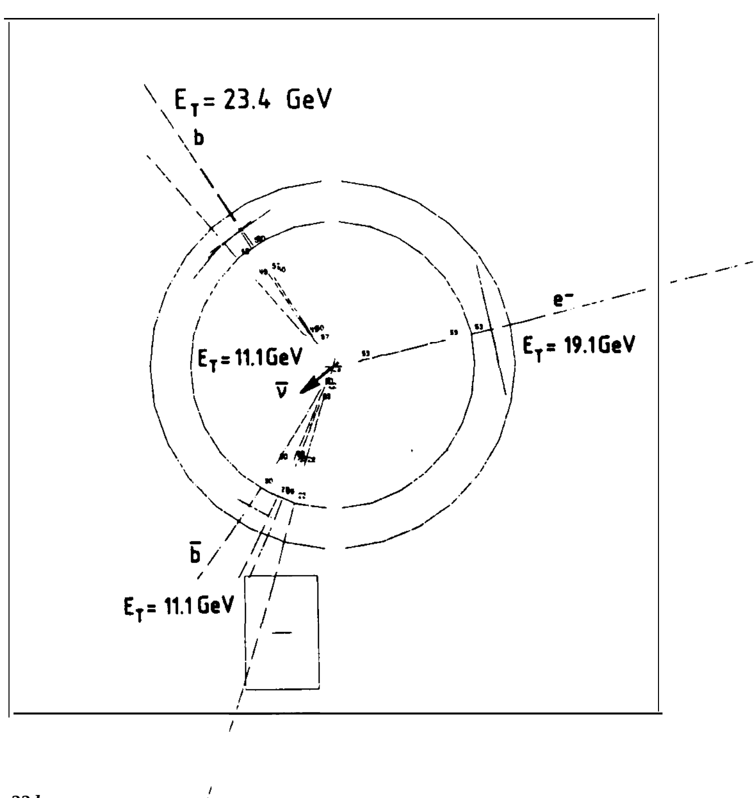

E_T = 23.4 GeV
b
e⁻
E_T = 19.1GeV
E_T = 11.1GeV
ν̄
b̄
E_T = 11.1 GeV

Fig. 22d.

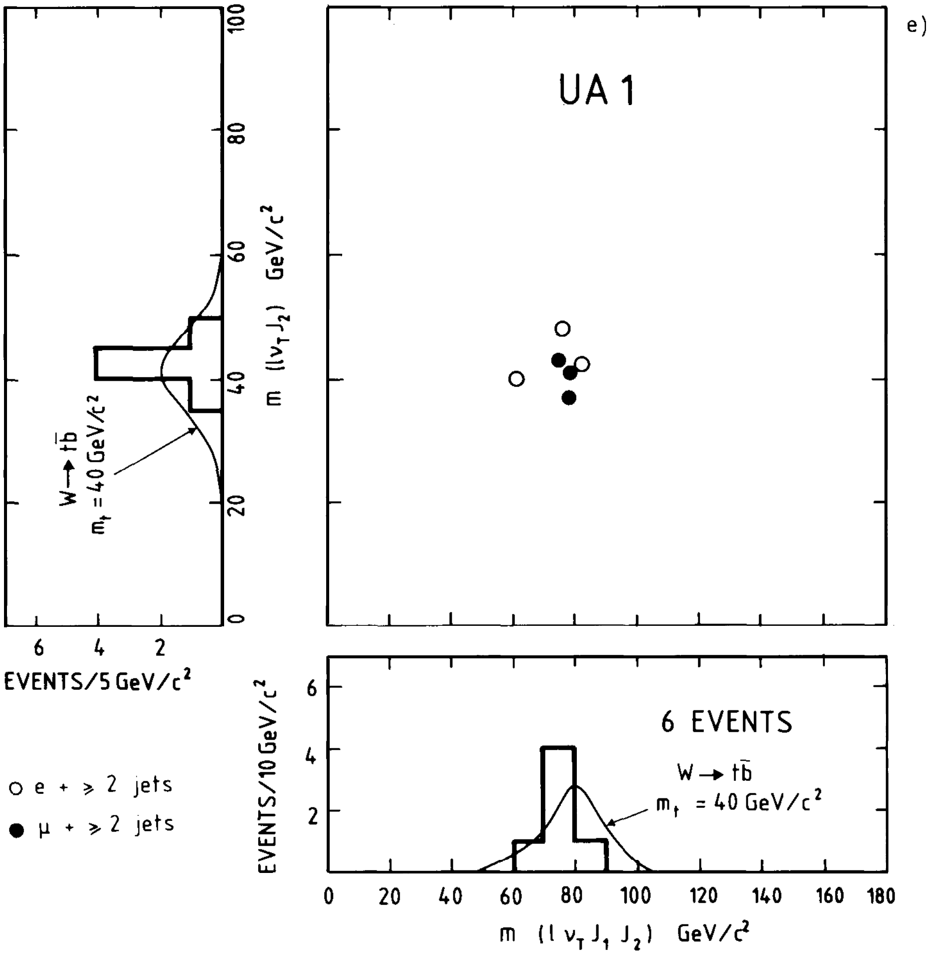


Fig. 22e.

that we now searched for electron and muon pairs rather than for electron--neutrino coincidence. The process is then

$$\bar{p}+p \rightarrow Z^0+X, \quad Z^0 \rightarrow e^+ + e^- \text{ or } \mu^+\mu^-.$$

This reaction is approximately a factor of 10 less frequent than the corresponding $W^\pm$ leptonic decay channels. A few events of this type were therefore expected in our muon or electron samples. Evidence for the existence of the Z^0 in the range of masses accessible to the UAl experiment has also been derived from weak-electromagnetic interference experiments at the highest PETRA energies, where deviations from point-like expectations have been reported (Fig. 23).

We first looked at events of the type $Z^0 \rightarrow e^+e^-$ [25,26]. As in the case of the $W^\pm$ search, an electron signature was defined as a localized energy deposition in two contiguous cells of the electromagnetic detectors with Er>25 GeV, and a small (or no) energy deposition (≤800 MeV) in the

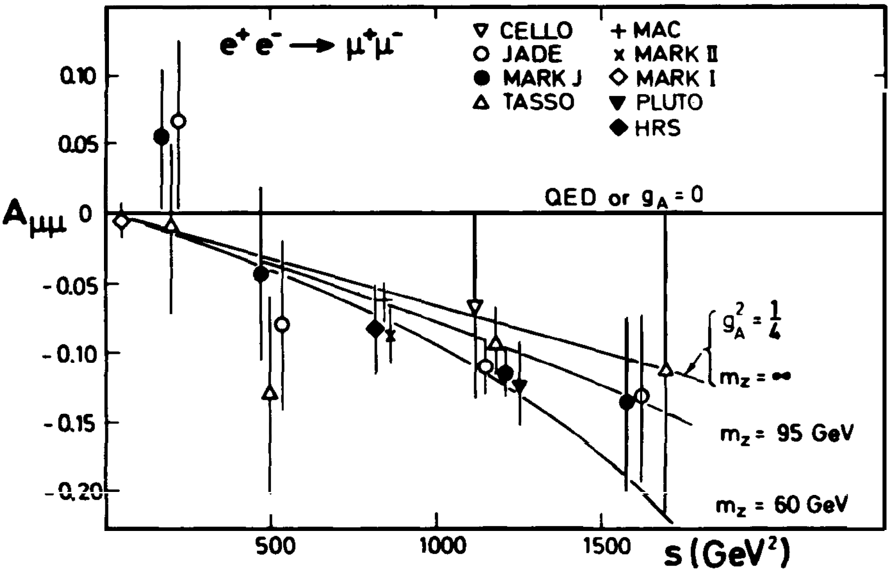


Fig. 23. Experimental evidence for a weak-electromagnetic interference effect in the process $e^+e^- \to \mu^+\mu^-$ at high-energy colliding beams. It can be seen that data are better fitted if the presence of a finite mass m_Z propagator is assumed.

hadron calorimeters immediately behind them. The isolation requirement was defined as the absence of charged tracks with momenta adding up to more than 3 GeV/c of transverse momentum and pointing towards the electron cluster cells. The effects of the successive cuts on the invariant electron-electron mass are shown in Fig. 24. Four e^+e^- events survived cuts, consistent with a common value of (e^+e^-) invariant mass. One of these events is shown in Figs. 25 and 26. As can be seen from the energy deposition plots (Fig. 27), the dominant feature of the four events is two very prominent electromagnetic energy depositions. All events appear to balance the visible total transverse energy components; namely, there is no evidence for the emission of energetic neutrinos. Except for the one track of event D which travels at less than 15” parallel to the magnetic field, all tracks are shown in Fig. 28, where the momenta measured in the central detector are compared with the energy deposition in the electromagnetic calorimeters. All tracks but one have consistent energy and momentum measurements. The negative track of event C shows a value of (9±1) GeV/c, much smaller than the corresponding deposition of (49±2) GeV. This event can be interpreted as the likely emission of a hard ‘photon’ accompanying the electron.

The same features are apparent also from the events in which a pair of muons [27] were emitted. A sharp peak (Fig. 29) is visible for high-mass dimuons. Within the statistical accuracy the events are incompatible with additional neutrino emission. They are all compatible with a common mass value:

$$\langle m_{\mu\mu} \rangle = 85.8^{+7.0}_{-5.4}\ \mathrm{GeV}/c^2,$$

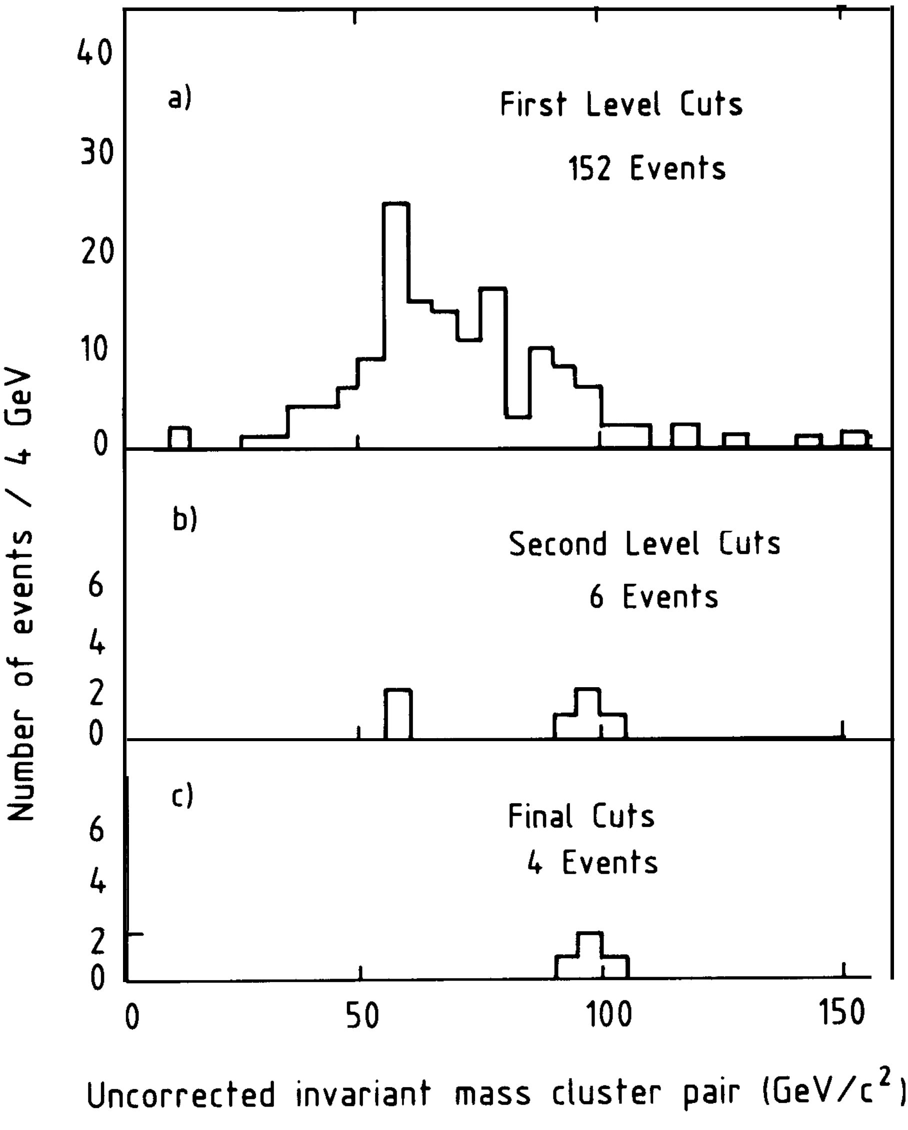


Fig. 24. **Invariant mass distribution (uncorrected) of two electromagnetic clusters: a) with E_T>25** GeV; b) as above, and a track with p_T>7 GeV/c and projection length of more than 1 cm pointing to the cluster. In addition, a small energy deposition in the hadron calorimeters immediately behind (<0.8 GeV) ensures the electron signature. Isolation is required with Σp_T<3 GeV/c for all other tracks pointing to the cluster. c) The second cluster also has an isolated track.

consistent with the value measured for $Z^0 \rightarrow e^+e^-$:

$$\langle m_{ee} \rangle = 95.6 \pm 1.4\ (2.9)\ \mathrm{GeV}/c^2,$$

where the first error accounts for the statistical error and the second for the uncertainty of the overall energy scale of the calorimeters. The average value for the nine Z^0 events found in the UAl experiment is $m_{Z^0} = 93.9 \pm 2.9$ GeV/c^2, where the error includes systematic uncertainties.

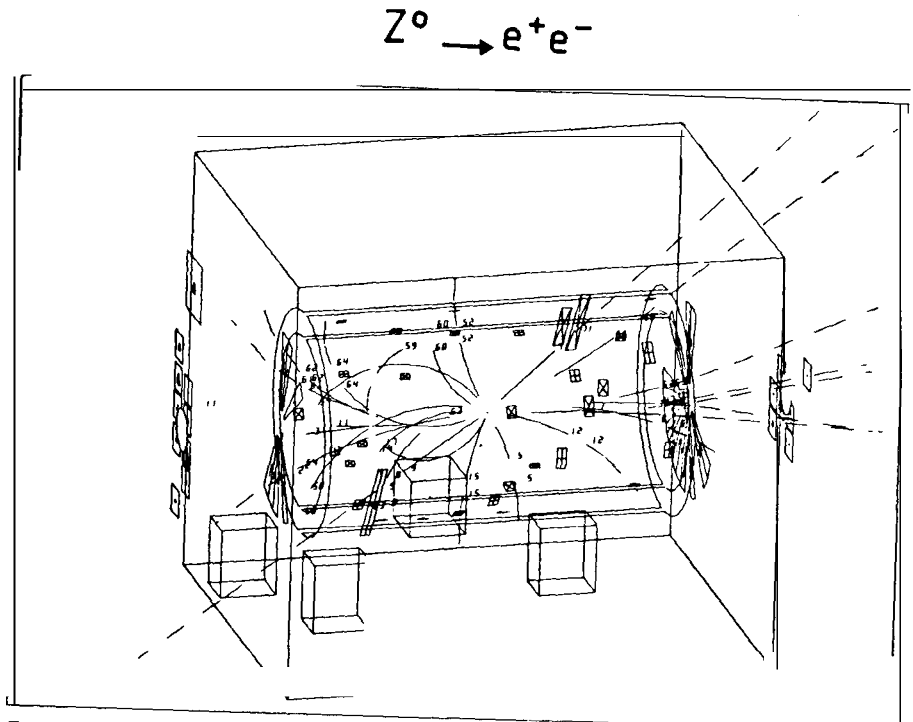


Fig. 25. Event display. All reconstructed vertex-associated tracks and all calorimeter hits are displayed.

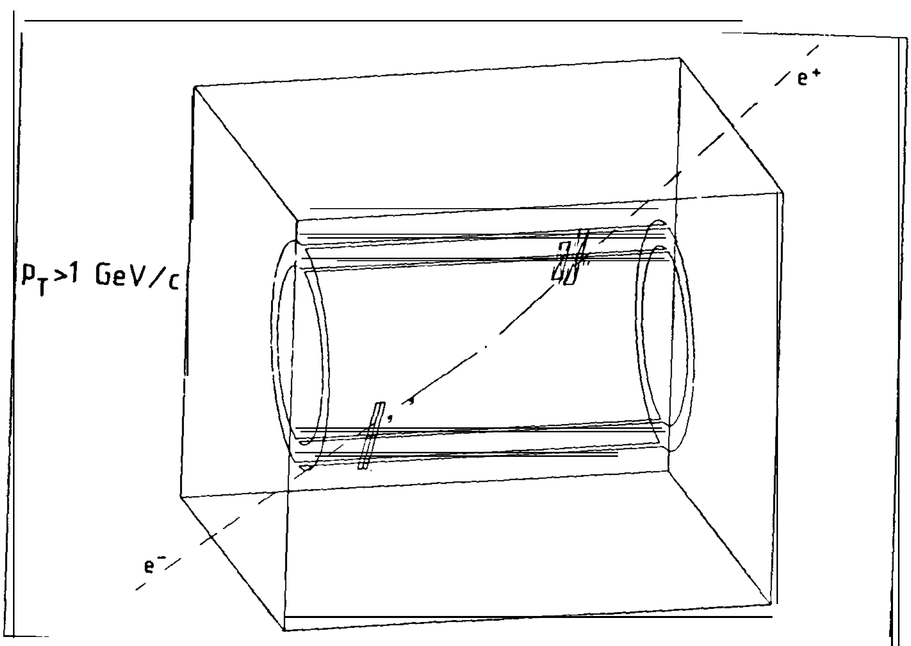


Fig. 26 The same as Fig. 25, but thresholds are raised to $p_T>2$ GeV/c for charged tracks and $E_T>2$ GeV for calorimeter hits. We remark that only the electron pair survives these mild cuts.

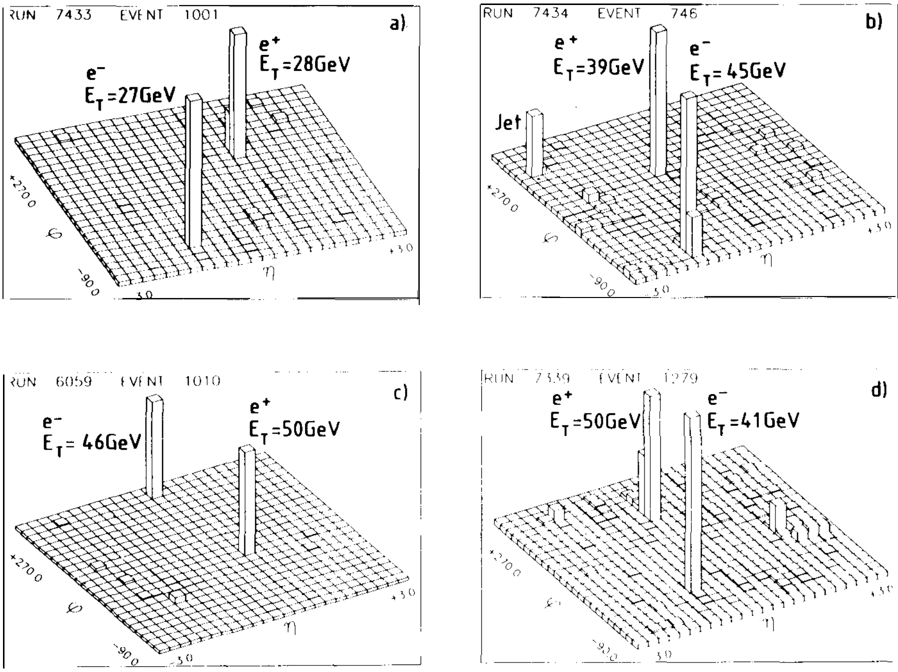


Fig. 27. Electromagnetic energy depositions at angles $>5°$ with respect to the beam direction for the four electron pairs.

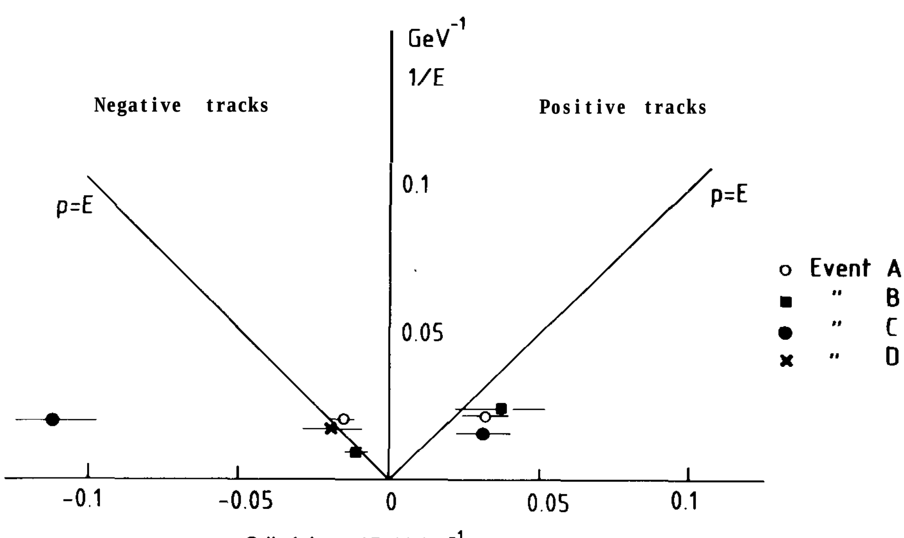


Fig. 28. Magnetic deflection in 1/p units compared with the inverse of the energy deposited in the electromagnetic calorimeters. Ideally, all electrons should lie on the l/E= 1/p line.

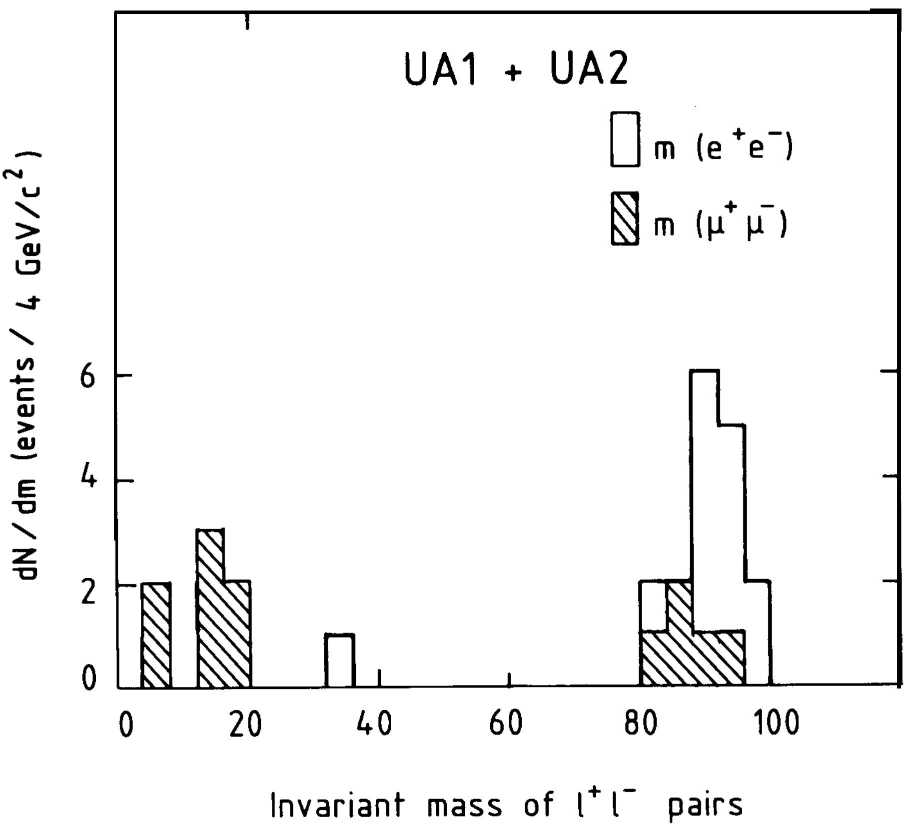


Fig. 29. Invariant mass distribution of dilepton events from UAl and UA2 experiments. A clear peak is visible at a mass of about 95 GeV/c^2.

The integrated luminosity for the present data sample is 108 nb^{-1}, with an estimated uncertainty of 15 %. With the geometrical acceptance of 0.37, the cross-section, calculated using the four events, is

$$(\sigma \cdot B)_{\mu\mu} = 100 \pm 50 (\pm 15) \text{ pb},$$

where the last error includes the systematics from the acceptance and from the luminosity. This value is in good agreement both with Standard Model predictions [22] and with our results for $Z^0 + e^+e^-$, namely $(\sigma \cdot B)_{ee} = 41 \pm 21 (\pm 7)$ pb. From the electron and the muon channels we obtain the average cross-section of

$$(\sigma \cdot B)_{\ell\ell} = 58 \pm 21 (\pm 9) \text{ pb}.$$

11. *Comparing theory with experiment*

The experiments discussed in the previous section have shown that the W particle has most of the properties required in order to be the carrier of weak interactions. The presence of a narrow dilepton peak has been seen around 95 GeV/c^2. Rates and features of the events are consistent with the hypothesis

Table 2. $W^{\pm}$ and Z^0 parameters from the UAl and UA2 experiments

	UAI	UA2
N(W→ eν)	52”	37^b
m_w (GeV/c^2)	80.9±1.5±2.4	83.1±1.9±1.3
Γ_W (90 % CL)	≤7 GeV	–
(σB) (nb)	0.53±0.08±0.09	0.53±0.10±0.10
N(W→ μν)	14	–
m_w (GeV/c^2)	81.0^{+6}_{-7}	–
(σB) (nb)	0.67±0.17±0.15	–
N (Z^0+ e^+e^-)	$3+1^c$	$7+1^c$
m_{z^0} (GeV/c^2)	95.6±1.4±2.9	92.7±1.7±1.4
Γ_{Z^0} (90 % CL)	S8.5 GeV	S6.5 GeV
(σB) (nb)	0.05±0.02±0.009	0.11+0.04+0.02
$N(Z^0 \to \mu^+\mu^-)$	$4+1^c$	–
m_{Z^0} (GeV/c^2)	85.6±6.3	
(σB) (nb)	0.105+0.05+0.15	–
$\sin^2\theta_w = 38.5/m_W$	0.226±0.015	0.216±0.010±0.007
$\varrho = [m_w/m_z \cos\theta_w]^2$	0.968±0.045	1.02±0.06

[a] $p_T^\nu >$ 15 GeV/c
[b] $p_T^e > 25$ GeV/c
[c] $Z^0 \to \ell^+\ell^-\gamma$ (E_γ > 20 GeV)

that the neutral partner of the $W^{\pm}$ has indeed been observed. At present the statistics are not sufficient to test the form of the interaction experimentally; neither has parity violation been detected. However, the precise values of the masses of Z^0 and $W^{\pm}$ now available constitute a critical test of the idea of unification between weak and electromagnetic forces, and in particular of the predictions of the SU(2) x U(1) theory of Glashow, Weinberg and Salam [6]. A careful account of systematic errors is needed in order to evaluate an average between the mass determination for the two collider experiments, UA1 and UA2 [28]. Table 2 summarizes all experimental information related to $W^{\pm}$ and Z^0.

The charged vector boson mass is

$$m_{W^\pm} = (80.9 \pm 1.5)\ \mathrm{GeV}/c^2 \quad \text{(statistical errors only)},$$

to which a 3 % energy scale uncertainty must be added. In this report a value for the Z^0 mass, $m_{Z^0} = (95.1 \pm 2.5)$ GeV/c^2, has been given. Neglecting systematic errors, a mass value is found with somewhat smaller errors:

$$m_{Z^0} = (95.6 \pm 1.4)\ \mathrm{GeV}/c^2 \quad \text{(statistical errors only)},$$

to which the same scale uncertainty as that for the $W^{\pm}$ applies. The quoted errors include: i) the neutral width of the Z^0 peak, which is found to be $\Gamma < 8.5$ GeV/c^2 (90 % confidence level); ii) the experimental resolution of counters; and iii) the r.m.s. spread between calibration constants of individual elements.

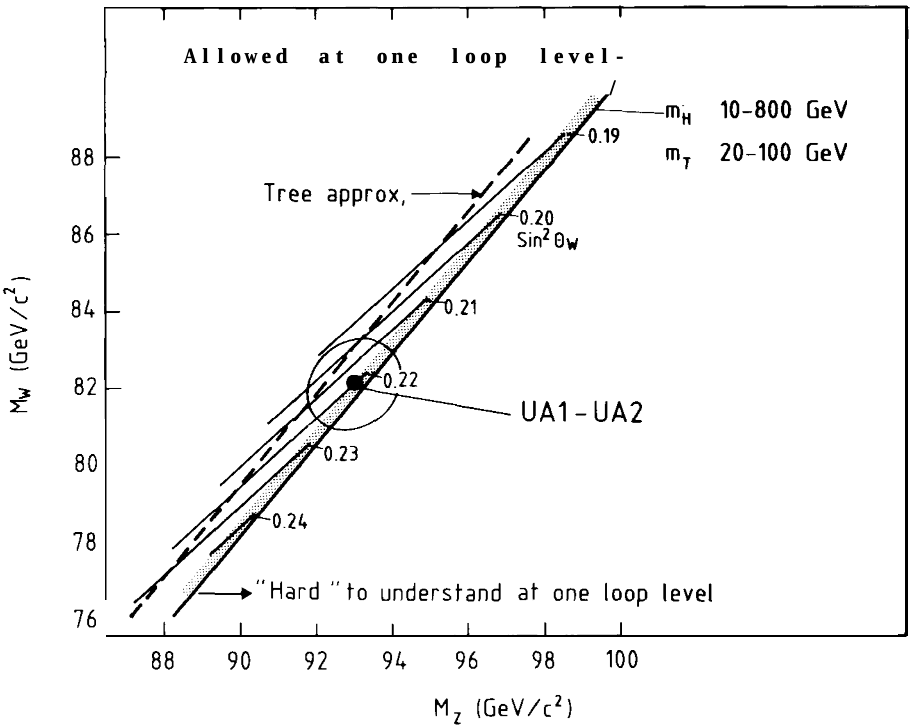


Fig. 30. Comparison between the Standard Model and the experimental results (UA1 and UA2 combined). Theory is from Ref. [29].

It should be remarked that the masses of the IVBs have the following prediction:

$$m_W = \left[\pi\alpha/\sqrt{2}\, G_F \sin^2\theta_W (1-\Delta r)\right]^{1/2},$$

$$m_Z = m_W/\cos\theta_W,$$

where the value Ar represents the effect of the higher-order radiative corrections, and the second equation can be used as a definition of the Weinberg angle θ_W. Since G_F and a are known, θ_W can be eliminated between equations:

$$m_Z = m_W/(1 - A^2/m_W^2)^{1/2},$$

$$\Delta r = A^2 m_Z^2 / \left[m_W^2 (m_W + m_Z)(m_Z - m_W)\right],$$

$$A = (37.2810 \pm 0.0003)\ \text{GeV}.$$

Radiative corrections are quite large [29] and detectable at the present level of accuracy. Calculations of order $O(\alpha) + O(\alpha^2 \ln m)$ give the following result:

$$\text{Ar} = 0.0696 \pm 0.0020,$$

which is insensitive to the parameters

$$\sin^2\theta_W = 0.217,$$

$$m_t = 40 \text{ GeV}/c^2, \quad m_b = 5 \text{ GeV}/c^2.$$

The main effect can be understood as α being a running coupling constant, namely:

$$a = 1/137.035962, \quad \text{at } q^2 = 0,$$

$$a = 1/137.5, \quad \text{at } q^2 = m_W^2.$$

In Fig. 30 we have plotted m_Z against m_W. The elliptical shape of the errors reflects the uncertainty in the energy scale. It can be seen that there is excellent agreement with the expectations of the SU(2)xU(1) Standard Model [29].

We can then extract the renormalized value of $\sin^2\theta_W$ at mass scale m_W. Inserting the value of m_W one finds

$$\sin^2\theta_W = 0.220\pm0.009,$$

In excellent agreement with the renormalized value of $\sin^2\theta_W=0.215\pm0.014$ deduced from neutral-current experiments. Using the information of the Z^0 mass, one can determine the parameter ϱ, related immediately to the isospin of the Higgs particle:

$$\varrho = m_W^2/m_{Z^0}^2 \cos^2\theta_W.$$

Using the experimental values, one finds

$$\varrho = 1.000\pm0.036,$$

in perfect agreement with the prediction of $\varrho = 1$ for a Higgs doublet. Let us point out that 9 deviates from 1 at most by 3%, owing to radiative corrections involving possible new fermion generations. The present value seems to indicate no such new fermion families.

We conclude that, within errors, the observed experimental values are completely compatible with the SU(2)xU(1) model, thus supporting the hypothesis of a unified electroweak interaction.

ACKNOWLEDGEMENTS

This lecture is based on the work of the UAl Collaboration team, and I would like to express my appreciation of their remarkable achievements which have led to so many exciting results. At present the following persons are members of the collaboration: G. Arnison, A. Astbury, B. Aubert, C. Bacci, A. Bezaguet, R. K. Bock, T. J. V. Bowcock, M. Calvetti, P. Catz, P. Cennini, S. Centro, F. Ceradini, S. Cittolin, D. Cline, C. Cochet, J. Colas, M. Corden, D. Dallman,

D. Dau, M. DeBeer, M. Della Negra, M. Demoulin, D. Denegri, A. Diciaccio, D. DiBitonto, L. Dobrzynski, J. D. Dowell, K. Eggert, E. Eisenhandler, N. Ellis, P. Erhard, H. Faissner, M. Fincke, G. Fontaine, R. Frey, R. Frühwirth, J. Garvey, S. Geer, C. Ghesquiere, P. Ghez, K. L. Giboni, W. R. Gibson, Y. Giraud-Heraud, A. Givernaud, A. Gonidec, G. Grayer, T. Hansl-Kozanecka, W. J. Haynes, L. O. Hertzberger, C. Hodges, D. Hoffmann, H. Hoffmann, D. J. Holthuizen, R. J. Homer, A. Honma, W. Jank, G. Jorat, P. I. P. Kalmus, V. Karimaki, R. Keeler, I. Kenyon, A. Kernan, R. Kinnunen, W. Kozanecki, D. Kryn, F. Lacava, J. P. Laugier, J. P. Lees, H. Lehmann, R. Leuchs, A. Leveque, D. Linglin, E. Locci, J. J. Malosse, T. Markiewicz, G. Maurin, T. McMahon, J. P. Mendiburu, M. N. Minard, M. Mohammadi, M. Moricca, K. Morgan, H. Muirhead, F. Muller, A. K. Nandi, L. Naumann, A. Norton, A. Orkin-Lecourtois, L. Paoluzi, F. Pauss, G. Piano Mortari, E. Pietarinen, M. Pimiä, J. P. Porte, E. Radermacher, J. Ransdell, H. Reithler, J. P. Revo1, J. Rich, M. Rijssenbeek, C. Roberts, J. Rohlf, P. Rossi, C. Rubbia, B. Sadoulet, G. Sajot, G. Salvi, G. Salvini, J. Sass, J. Saudraix, A. Savoy-Navarro, D. Schinzel, W. Scott, T. P. Shah, D. Smith, M. Spiro, J. Strauss, J. Streets, K. Sumorok, F. Szoncso, C. Tao, G. Thompson, J. Timmer, E. Tscheslog, J. Tuominiemi, B. Van Eijk, J. P. Vialle, J. Vrana, V. Vuillemin, H. D. Wahl, P. Watkins, J. Wilson, R. Wilson, C. E. Wulz, Y. G. Xie, M. Yvert and E. Zurfluh.

REFERENCES

1. E. Fermi, Ric. Sci. 4 (2), 491 (1933), reprinted *in* E. Fermi, Collected Papers (eds. E. Segré et al.) (University of Chicago Press, Chicago, Ill., 1962), Vol. 1, p. 538; Z. Phys. 88, 161 (1934); English translation: F. L. Wilson, Am. J. Phys. 36 1150 (1968).
2. M. Gell-Mann and M. Levy, Nuovo Cimento 16, 705 (1960).
3. N. Cabibbo, Phys. Rev. Lett. 10, 531 (1963).
   M. Kobayashi and K. Maskawa, Progr. Theor. Phys. 49, 652 (1973).
4. S. S. Gerschtein and J. R. Zeldovich, Z. Eksp. Teor. Fiz. 29, 698 (1955).
   R. P. Feynman and M. Gell-Mann, Phys. Rev. 109, 193 (1958).
   E. C. G. Sudarshan and R. E. Marshak, Phys. Rev. 109, 1860 (1958).
   J. J. Sakurai, Nuovo Cimento 7, 649 (1958).
5. O. Klein, in Proc. Symp. on Les Nouvelles Theories de la Physique, Warsaw, 1938 (Institut International de Coopération Intellectuelle, Paris, 1939), p.6.
   O. Klein, Nature 161, 897 (1948).
6. S. L. Glashow, Nucl. Phys. 22, 579 (1961).
   S. Weinberg, Phys. Rev. Lett. 29, 1264 (1967).
   A. Salam, Proc. 8th Nobel Symposium (ed. N. Svartholm) (Almqvist and Wiksell, Stockholm, 1968), p. 367.
7. T. D. Lee and C. N. Yang, Phys. Rev. 119, 1410 (1960).
8. The first realistic scheme for colliding beams was discussed by D. W. Kerst et al., Phys. Rev. 202, 590 (1956). The first suggestion for proton-antiproton intersecting beams was given by G. I. Budker, Proc. Int. Symp. on Electron and Positron Storage Rings, Saclay, 1966 (eds. H. Zyngier and E. Crémieu-Alcan) (PUF, Paris, 1966), p. II-l-l. G. I. Budker, At. Energ. 22, 346 (1967).
9. C. Rubbia, P. McIntyre and D. Cline, Proc. Int. Neutrino Conf., Aachen, 1976 (eds. H. Faissner, H. Reithler and P. Zerwas) (Vieweg, Braunschweig, 1977), p. 683.
10. The 300 GeV programme, CERN/1050 (14 January 1972).

11. S. van der Meer, Internal Report CERN ISR-PO/72-31 (1972).
D. Möhl, G. Petrucci, L. Thorndahl and S. van der Meer, Phys. Rep. 58, 73 (1980).
12. Design study of intersecting storage rings (ISR) for the CERN Proton Synchrotron, CERN AR/Int. SG/64-9 (1964).
13. A. Astbury et al., A 4x solid-angle detector for the SPS used as a proton-antiproton collider at a centre-of-mass energy of 540 GeV, Proposal, CERN-SPSC 78-6/P92 (1978).
14. The UAl Collaboration is preparing a large and comprehensive report on the detector (1984). For details see:
M. Barranco Luque et al., Nucl. Instrum. Methods 176, 175 (1980).
M. Calvetti et al., Nucl. Instrum. Methods 176, 255 (1980).
M. Calvetti et al., Proc. Int. Conf. on Instrumentation for Colliding Beam Physics, Stanford, 1982 (SLAC-250, Stanford, 1982), p. 16.
M. Calvetti et al., IEEE Trans. Nucl. Sci. NS-30, 71 (1983).
J. Timmer, The UAl detector, Antiproton proton physics and the W discovery, presented at the 3rd Moriond Workshop, La Plagne, France, March 1983 (Ed. Tran Thanh Van) (Editions Frontierès, Gif-sur-Yvette, France), p. 593.
E. Locci, Thèse de doctorat ès sciences, Paris, 1984, unpublished.
M. J. Corden et al., Phys. Scr. 25, 5 and 11 (1982).
M. J, Corden et al., Rutherford preprint RL-83-116 (1983).
K. Eggert et al., Nucl. Instrum. Methods 176, 217 (1980).
J. C. Santiard, CERN EP Internal Report 80-04 (1980).
K. Eggert et al., Nucl. Instrum. Methods ***176,*** 223 (1980) and 188, 463 (1981).
G. Arnison et al., Phys. Lett. ***121B 77*** (1983).
G. Arnison et al., Phys. Lett. ***128B 336*** (1983).
The UA2 Collaboration have described their detector, which however makes a more limited use of the missing-energy concept, in Proc. 2[nd] Int. Conf. on Physics in Collisions, Stockholm, 1982 (Plenum Press, New York, 1983), p. 67.
A. G. Clark, Proc. Int. Conf. on Instrumentation for Colliding Beam Physics, Stanford, 1982 (SLAG-250, Stanford, 1982), p. 169.
B. Mansoulié, The UA2 apparatus at the CERN $p\bar{p}$ Collider, Antiproton proton physics and the W discovery, presented at the 3rd Moriond Workshop, La Plagne, France, March 1983 (Ed. Tran Thanh Van) (Editions Frontières, Gif-sur-Yvette, France), p. 609.
15. G. Arnison et al., Phys. Lett. ***122B,*** 103 (1983).
16. M. Banner et al., Phys. Lett. ***122B, 476*** (1983).
17. G. Arnison et al., Phys. Lett. ***129B, 273*** (1983).
C. Rubbia, Proc. Int. Europhysics Conf. on High-Energy Physics, Brighton, 1983 (eds. J. Guy and C. Costain) (Rutherford Appleton Lab., Didcot, 1983), p. 860.
18. G. Arnison et al., Phys. Lett. ***139B,*** 115 (1984).
19. C. Rubbia, Physics results of the UAl Collaboration at the CERN proton-antiproton collider, Proc. Int. Conf. on Neutrino Physics and Astrophysics, Dortmund, 1984 (ed. K. Kleinknecht), p. 1.
20. J, F. Owens and E. Reya, Phys. Rev. ***D17, 3003*** (1978).
F. Halzen and W. Scott, Phys. Lett. 78B, 318 (1978).
F. Halzen, A. D. Martin and D. M. Scott, Phys. Rev. ***D25, 754*** (1982).
P. Aurenche and R. Kinnunen, Annecy preprint LAPP-TH-78 (1983).
A. Nakamura, G. Pancheri and Y. Srivastava, Frascati preprint LNF-83-44 (1983).
21. G. Altarelli, R. K. Ellis, M. Greco and G. Martinelli, Nucl. Phys. ***B246,*** 12 (1984).
G. Altarelli, G. Parisi and R. Petronzio, Phys. Lett. ***76B,*** 351 and 356 (1978).
22. F. E. Paige, in Proc. Topical Workshop on the Production of New Particles in Super High Energy Collisions, Madison, 1979 (eds. V. Barger and F. Halzen) (AIP, New York, 1979).
F. E. Paige and S. D. Protopopescu, ISAJET program, BNL 29777 (1981).
G. Altarelli, R. K. Ellis and G. Martinelli, Nucl. Phys. ***Bl43,*** 521 (1978); (E) ***B196, 544*** (1978); ***B157,*** 461 (1979).
J. Kubar-Andre and F. E. Paige, Phys. Rev. ***D19,*** 221 (1979).

23. M. Jacob, Nuovo Cimento 9, 826 (1958).
M. Jacob, unpublished (see, for instance, in C. Rubbia, Proc. 9^{th} Topical Conf. on Particle Physics, Honolulu, 1983).
24. G. Arnison et al., Phys. Lett. ***147B, 493*** (1984).
25. G. Arnison et al., Phys. Lett. ***126B, 398*** (1983).
26. P. Bagnaia et al., Phys. Lett. ***2298,*** 150 (1983).
27. G. Arnison et al., Phys. Lett. ***147B,*** 241 (1984).
28. C. Rubbia, Proc. Int. Europhysics Conf. on High-Energy Physics, Brighton, 1983 (eds. J. Guy and C. Costain) (Rutherford Appleton Lab., Didcot, 1983), p. 860.
29. M. Consoli, Proc. Third Topical Workshop on $p\bar{p}$ Collider Physics, Rome, 1983 (eds. C. Bacci and G. Salvini) (CERN 83-04, Geneva, 1983), p. 478.
M. Consoli, S. L. Presti and L. Maiani, Univ. Catania preprint PP/738-14/1/1983.
M. Veltman, Proc. Int. Europhysics Conf. on High-Energy Physics, Brighton, 1983 (eds. J. Guy and C. Costain) (Rutherford Appleton Lab., Didcot, 1983), p. 880.
W. Marciano, BNL preprint 33819 (1983).

Simon van der Meer

SIMON VAN DER MEER

I was born in 1925, in The Hague, the Netherlands, as the third child of Pieter van der Meer and Jetske Groeneveld, both of Frisian origin. I had three sisters.

My father was a schoolteacher and my mother came from a teacher's family. Under these conditions it is not astonishing that learning was highly prized; in fact, my parents made sacrifices to be able to give their children a good education.

I visited the Gymnasium in The Hague and passed my final examination (in the sciences section) in 1943. Because the Dutch universities had just been closed at that time under the German occupation, I spent the next two years attending the humanities section of the Gymnasium. Meanwhile, my interest in physics and technology had been growing; I dabbled in electronics, equipped the parental home with various gadgets and assisted my brilliant and inspiring physics teacher (U. Ph. Lely) with the preparation of numerous demonstrations.

From 1945 onwards, I studied "Technical Physics" at the University of Technology, Delft, where I specialized in measurement and regulation technology under C. J. D. M. Verhagen. The physics taught in this newly created subsection of an old and established engineering school, although of excellent quality, was of necessity somewhat restricted and I have often felt regrets at not having had the intensive physics training that many of my colleagues enjoyed. Nevertheless, if I have at times been able to make original contributions in the accelerator field, I cannot help feeling that to a certain extent my slightly amateur approach in physics, combined with much practical experience, was an asset.

After obtaining my engineering degree in 1952, I worked in the Philips Research Laboratory, Eindhoven, mainly on high-voltage equipment and electronics for electron microscopes. In 1956 I moved to Geneva to join the recently founded European Organization for Nuclear Research (CERN), where I have been working ever since on many different projects, in an agreeable and stimulating international atmosphere.

To start with, my work (under the leadership of J. B. Adams and C. A. Ramm) was concerned mainly with technical design: poleface windings, multipole correction lenses for the 28 GeV synchrotron and their power supplies. My interest in matters more directly concerned with the handling of particles was growing, in the meantime, stimulated by many contacts with people understanding accelerators. After working for a year on a separated antiproton beam (1960), I proposed a high-current, pulsed focusing device ("horn") aimed at increasing the intensity of a beam of neutrinos, then at the centre of interest at

CERN and elsewhere. The design of this monster, together with the associated neutrino flux calculations kept me busy until 1965, when I joined a small group, led by F. J. M. Farley, preparing the second "g-2" experiment for measuring the anomalous magnetic moment of the muon. I designed the small storage ring used and participated at all stages of the experiment proper, including part of the data treatment. This was an invaluable experience; not only did I learn the principles of accelerator design, but I also got acquainted with the lifestyle and way of thinking of experimental high-energy physicists.

From 1967 to 1976 I returned to more technical work when I was responsible for the magnet power supplies, first of the Intersecting Storage Rings (ISR) and then of the 400 GeV synchrotron (SPS). I kept up with accelerator ideas, however, and worked (during my ISR period) on a method for the luminosity calibration of storage rings and on stochastic cooling. The latter was, of course, aimed at increasing the ISR luminosity, but practical application seemed difficult at the time, mainly because the high beam intensity in the ISR would have made the cooling very slow. After developing a primitive theory (1968) I therefore did not pursue this subject. However, the work was taken up by others and in 1974 the first experiments were done in the ISR.

In 1976, Cline, McIntyre, Mills, and Rubbia proposed to use the SPS or the Fermilab ring as a pp collider. Accumulation of the needed antiprotons would clearly require cooling. At this time, my work on the SPS power supplies had just come to an end; I joined a study group on the $p\bar{p}$ project and an experimental team studying cooling in a small ring (ICE). The successful experiments in this ring and the work by Sacherer on theory and by Thorndahl on filter cooling showed that $\bar{p}$ accumulation by stochastic stacking was feasible. The collider project was approved and I became joint project leader with R. Billinge for the accumulator construction. Since then, I have worked with the group that commissioned and improved the ring and that is now preparing the construction of a second ring to increase the $\bar{p}$ stacking rate by an order of magnitude. As a spin-off from this work, I proposed the stochastic extraction method that is now used (in a much improved form) in the Low-Energy Antiproton Ring (LEAR).

In the meantime, in 1966, while skiing with friends in the Swiss mountains, I met my wife-to-be Catharina M. Koopman and after a very brief interval we decided to marry. This was certainly one of the best decisions I ever made; my life has since been far more interesting and colourful. We have two children: Esther (1968) and Mathijs (1970).

(added in 1991): In 1990 I retired from CERN.

Honours

Loeb Lecturer, Harvard University, 1981.

Duddell Metal, Institute of Physics, 1982.

Honorary Degree, Geneva University, 1983.

Honorary Degree, Amsterdam University, 1984.

Foreign Honorary Member, American Academy of Arts and Sciences, 1984.

Correspondent, Royal Netherlands Academy of Sciences, 1984.

STOCHASTIC COOLING AND THE ACCUMULATION OF ANTIPROTONS

Nobel lecture, 8 December, 1984

by

SIMON VAN DER MEER
CERN, CH- 1211 Geneva 23, Switzerland

1. A general outline of the $p\bar{p}$ project

The large project mentioned in the motivation of this year's Nobel award in physics includes in addition to the experiments proper described by C. Rubbia, the complex machinery for colliding high-energy protons and antiprotons (Fig. 1). Protons are accelerated to 26 GeV/c in the PS machine and are used to produce p's in a copper target. An accumulator ring (AA) accepts a batch of these with momenta around 3.5 GeV/c every 2.4 s. After typically a day of accumulation, a large number of the accumulated $\bar{p}$'s (~10^{11}) are extracted from the AA, reinjected into the PS, accelerated to 26 Gev/c and transferred to the large (2.2 km diameter) SPS ring. Just before, 26 Gev/c protons, also from the PS, have been injected in the opposite direction. Protons and antiprotons are then accelerated to high energy (270 or 310 Gev) and remain stored for many hours. They are bunched (in 3 bunches of about 4 ns duration each) so that collisions take place in six well-defined points around the SPS ring, in two of which experiments are located. The process is of a complexity that could only be mastered by the effort and devotion of several hundreds of people. Only a small part of it can be covered in this lecture, and I have chosen to speak about stochastic cooling, a method that is used to accumulate the antiprotons, and with which I have been closely associated.

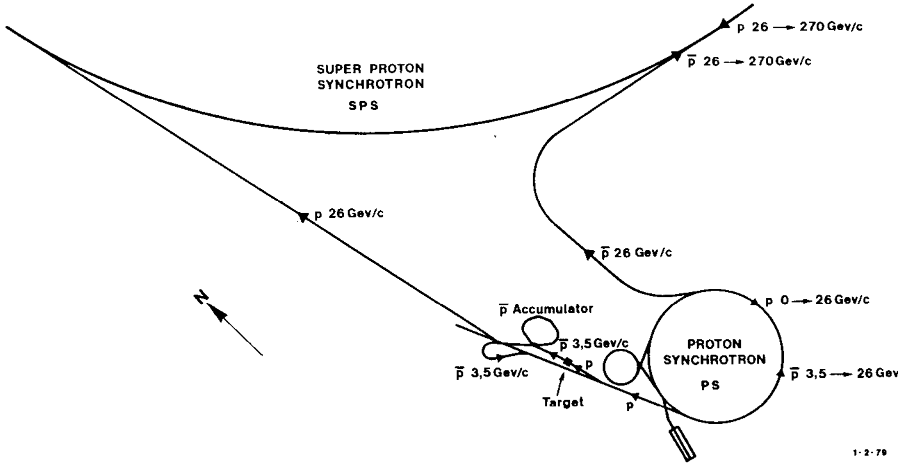


Fig. *1.* Overall layout of the $p\bar{p}$ project.

2. Cooling, why and how?

A central notion in accelerator physics is phase space, well-known from other areas of physics. An accelerator or storage ring has an acceptance that is defined in terms of phase volume. The antiproton accumulator must catch many antiprotons coming from the target and therefore has a large acceptance; much larger than the SPS ring where the p's are finally stored. The phase volume must therefore be reduced and the particle density in phase space increased. On top of this, a large density increase is needed because of the requirement to accumulate many $\bar{p}$ batches. In fact, the density in 6-dimensional phase space is boosted by a factor 10^9 in the AA machine.

This seems to violate Liouville's theorem that forbids any compression of phase volume by conservative forces such as the electromagnetic fields that are used by accelerator builders. In fact, all that can be done in treating particle beams is to distort the phase volume without changing the density anywhere.

Fortunately, there is a trick - and it consists of using the fact that particles are points in phase space with empty space in between. We may push each particle towards the centre of the distribution, squeezing the empty space outwards. The small-scale density is strictly conserved, but in a macroscopic sense the particle density increases. This process is called cooling because it reduces the movements of the particles with respect to each other.

Of course, we can only do this if we have information about the individual particle's position in phase space and if we can direct the pushing action against the individual particles. Without these two prerequisites, there would be no reason why particles rather than empty space would be pushed inwards. A stochastic cooling system therefore consists of a sensor (pick-up) that acquires electrical signals from the particles, and a so-called kicker that pushes the particles and that is excited by the amplified pick-up signals.

Such a system resembles Maxwell's demon, which is supposed to reduce the entropy of a gas by going through a very similar routine, violating the second law of thermodynamics in the process. It has been shown by Szilard[1] that the measurement performed by the demon implies an entropy increase that compensates any reduction of entropy in the gas. Moreover, in practical stochastic cooling systems, the kicker action is far from reversible; such systems are therefore even less devilish than the demon itself.

3. Qualitative description of betatron cooling

The cooling of a single particle circulating in a ring is particularly simple. Fig. 2 shows how it is done in the horizontal plane. (Horizontal, vertical and longitudinal cooling arc usually decoupled.)

Under the influence of the focusing fields the particle executes betatron oscillations around its central orbit. At each passage of the particle a so-called differential pick-up provides a short pulse signal that is proportional to the distance of the particle from the central orbit. This is amplified and applied to the kicker, which will deflect the particle. If the distance between pick-up and kicker contains an odd number of quarter betatron wavelengths and if the gain is chosen correctly, any oscillation will be cancelled. The signal should arrive at

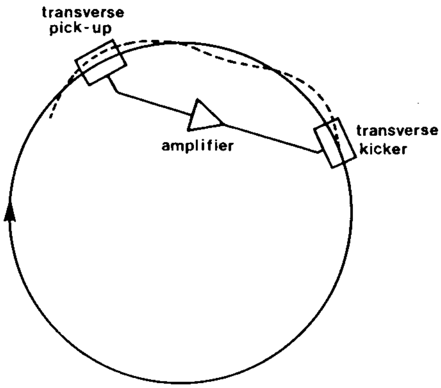


Fig. 2. Cooling of the horizontal betatron oscillation of a single particle

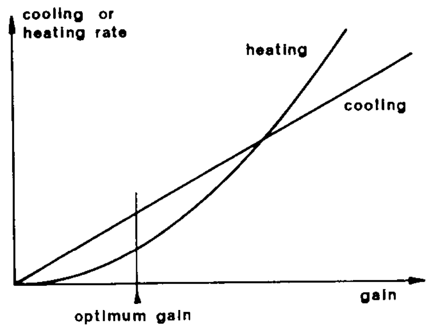


Fig. 3. Variation with system gain of the coherent cooling and incoherent heating effect

the kicker at the same time as the particle; because of delays in the cabling and amplifiers, the signal path must cut off a bend in the particle's trajectory.

In practice, there will not just be one particle, but a very large number (e.g. 10^6 or 10^{12}). It is clear that even with the fastest electronics their signals will overlap. Nevertheless, each particle's individual signal will still be there and take care of the cooling. However, we must now reduce the gain of the system because all the other particles whose signals overlap within one system response time will have a perturbing (heating) effect, as they will in general have a random phase with respect to each other. Fortunately, the perturbing effect is on average zero and it is only its second-order term that heats (i.e. increases the mean square of the amplitude). This is proportional to the square of the gain, whereas the cooling effect-each particle acting on itself-varies linearly with gain. As illustrated in Fig. 3, we may always choose the gain so that the cooling effect predominates.

4. Simplified analysis of transverse cooling

We shall now analyse the process sketched above in a somewhat approximative way, neglecting several effects that will be outlined later. The purpose is to get some feeling about the possibilities without obscuring the picture by too much detail.

In the first place, we shall assume a system with constant gain over a bandwidth W and zero gain outside this band. A signal passed by such a system may be described completely in terms of 2W samples per unit time. If we have N particles in the ring and their revolution time is T, each sample will on average contain

$$N_s = N/2WT \tag{1}$$

particles. We may now consider the system from two viewpoints:

a) we may look at each individual particle and combine the cooling by its own signal with the heating by the other particles,

b) we may look at the samples as defined above and treat each sample as the single particle of Fig. 1; this is justified because the samples are just resolved by the system.

The two descriptions are equivalent and yield the same result. For the moment, we shall adopt b). Incidentally, the name "stochastic cooling" originated' because from this viewpoint we treat a stochastic signal from random samples. However, viewpoint a) is more fundamental; cooling is not a stochastic process.

The pick-up detects the average position of each sample $\bar{x}$ and the gain will be adjusted so that this is reduced to zero, so that for each particle x is changed into $x-\bar{x}$. Averaging over many random samples, we see that the mean square x^2 is changed into

$$\overline{(x-\bar{x})^2} = \overline{x^2}-\bar{x}^2.$$

Therefore, the decrement of x^2 per turn is $\bar{x}^2/x^2 = 1/N_s$, and the cooling rate (expressed as the inverse of cooling time) is $1/t = 1/N_sT$. In fact, we have to divide this by four. One factor 2 occurs because the betatron oscillation is not always maximum at the pick-up as shown in Fig. 2. Both at the pick-up and at the kicker wc therefore lose by a factor equal to the sine of a random phase angle; the average of $\sin^2$ is 1/2. Another factor 2 is needed because it is usual to define cooling rate in terms of amplitude rather than its square. So we have, using (1)

$$\frac{1}{\tau} = \frac{1}{4N_sT} = \frac{W}{2N}\ . \tag{2}$$

This result, although approximative, shows that stochastic cooling is not a practical technique for proton accelerators; for a typical accelerator $N \approx 10^{13}$, so that even with a bandwidth of several GHz the cooling would be much too slow compared to the repetition rate. In storage rings, however, the available time is longer and sometimes the intensity is lower, so that the technique may become useful.

5. Mixing and thermal noise

In deriving the cooling rate, we assumed that all samples have a random population, without correlation between successive turns. The main reason why the sample populations change is the spread in energy between the particles, which results in a revolution frequency spread. The particles overtake each other, and if the spread of revolution time is large compared to the sample duration, we speak of "good mixing"; in this case the derivation above is valid. In practice, it is rarely possible to achieve this ideal situation. In particular with strongly relativistic particles a large spread of revolution frequency can only be obtained by a large spread in orbit diameter; for a given aperture this reduces the momentum spread that is accepted by the machine.

We may see how bad mixing influences the cooling by replacing the correction $\bar{x}$ in the derivation of the cooling rate by a smaller amount $g\bar{x}$. As a result we find in the same way

$$\frac{1}{\tau}=\frac{W}{2N}\left(2g-g^2\right). \tag{3}$$

Clearly, this is largest for $g = 1$.

It can be shown that the two terms correspond to the coherent, cooling effect (each particle cooled by its own signal) and the incoherent, heating effect from the other particles[3]. It is the second one that increases by bad mixing, because of the correlation between samples at successive turns. It may also increase if thermal noise is added to the signal (usually originating in the low-level amplifier attached to the pick-up). Thus, we may define a mixing factor M (= 1 for perfect mixing) and a thermal noise factor U (equal to the noise/ signal power ratio) and obtain

$$\frac{1}{\tau}=\frac{W}{2N}\left(2g-g^2(M+U)\right).$$

By optimizing g (now < 1) we find

$$\frac{1}{\tau}=\frac{W}{2N(M+U)} \tag{4}$$

6. Frequency domain analysis

This qualitative analysis may be made much more precise by considering the process from the frequency (instead of time) domain standpoint[4-5].

Each particle produces in the pick-up (considered to be ideal) a delta-function signal at each passage. For a sum pick-up, where the signal is independent of the transverse position, the Fourier transformation into the frequency domain results in a contribution at each harmonic of the revolution frequency (Fig. 4) while for a difference pick-up the modulation by the betatron oscillation splits up each line into two components[5]. For a collection of many particles with slightly different revolution frequencies, these lines spread out into bands, called Schottky bands because they represent the noise due to the finite number of charge carriers as described by Schottky[6].

The width of these bands increases towards higher frequency. The total power is the same for each band. The power density is therefore lower for the wider bands at high frequency up to the point where they start to overlap; beyond this point the bands merge and their combined density is constant with frequency. This is illustrated in Fig. 5 for so-called longitudinal lines (from a sum pick-up).

The cooling process may now be seen as follows. Firstly, each particle will cool itself with its own (coherent) signal. This means that at the frequency of each of its Schottky lines the phase of the corresponding sine-wave signal must be correct at the kicker, so that the latter exerts its influence in the right direction. Secondly, the other particles produce an incoherent heating effect at

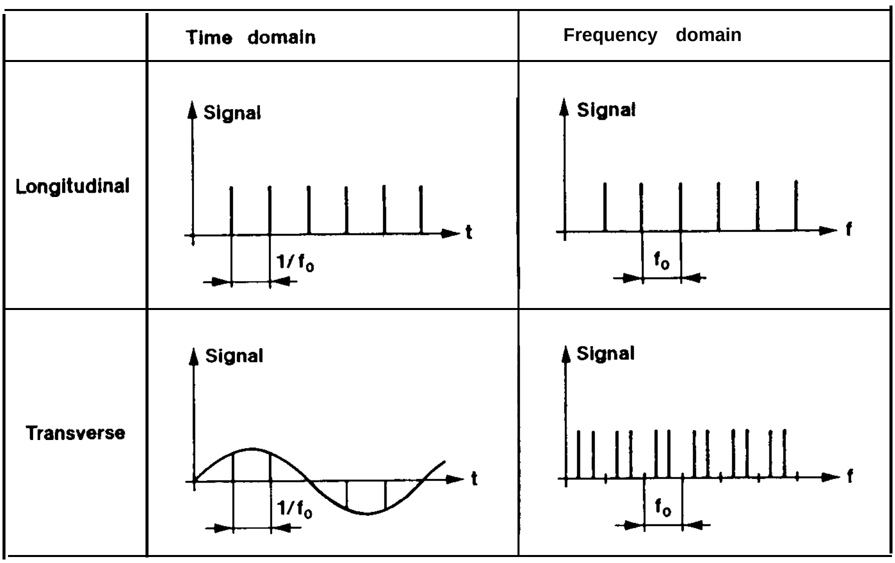


Fig. 4. Schottky signals in time domain and frequency domain

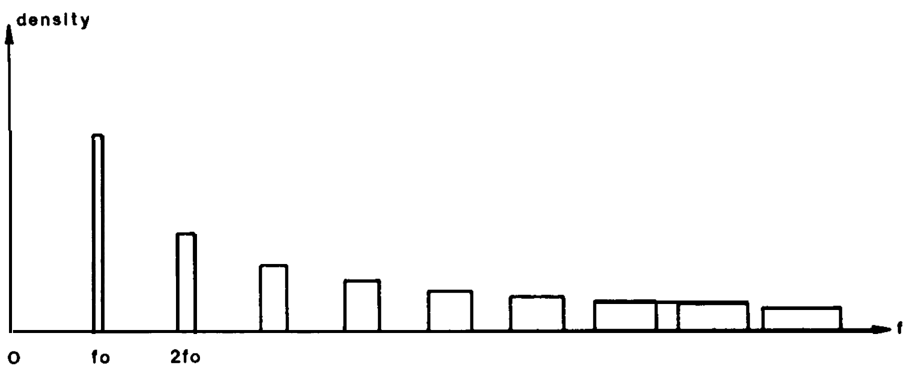


Fig. 5. Longitudinal Schottky bands originating from a large group of particles with slightly different revolution frequencies. At high frequencies the bands overlap.

each Schottky line proportional to the noise power density around that line[7]. Thus, only particles with frequencies very near to those of the perturbed particle will contribute. Any power density from thermal noise must of course be added to the Schottky power density.

For obtaining optimum cooling, the gain at each Schottky band should be adjusted so as to achieve an optimum balance between these two effects. If the bands are separated, the low-frequency ones have a higher density. This requires a lower gain and leads to less cooling for these bands. This is exactly the same effect that we called "bad mixing" in the time domain. At higher frequencies where the bands overlap we have good mixing and the gain should be independent of frequency.

Note that the picture given here (i.e. heating only caused by signals near the particle's Schottky frequencies) is completely different from the time-domain picture, where it seemed that particles in the same sample all contribute, independent of their exact revolution frequency. In fact, the latter is only true if the mixing is perfect and the samples arc statistically independent. In the more general case, it turns out that both the optimum gain and the optimum cooling rate per line are inversely proportional to the density dN/df around that line, rather than to the total number of particles N. In the time domain treatment this was expressed by the mixing factor M, but the dependence of the parameters on frequency was lost.

There is yet another mixing effect that we have neglected so far. While moving from the pick-up to the kicker, each sample will already mix to a certain extent with its neighbours. This harmful effect may be described in the frequency domain as a phase lag increasing with frequency (particles with higher revolution frequency arrive too early at the kicker, so that their signal is too late). It appears quite difficult to correct this by means of filters at each Schottky band; on the other hand, in practical cases the effect is usually not very serious[8].

7. Beam feedback

Another aspect that we have not yet considered is essential for the correct analysis of a cooling system. This is the feedback loop formed by the cooling chain together with the beam response (Fig. 6). Any signal on the kicker will

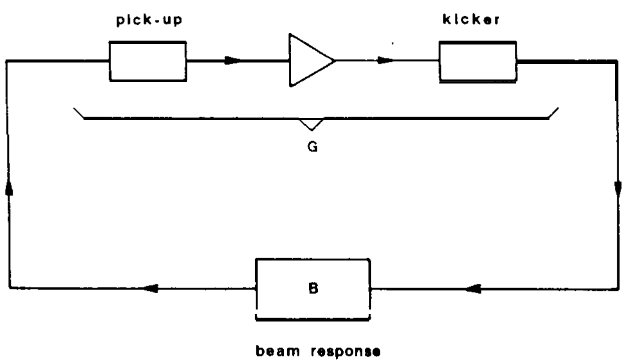


Fig. 6. Beam feedback effect. The loop is closed by the coherent beam response B

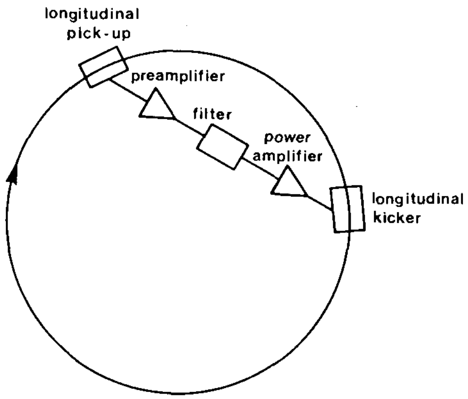


Fig. 7. Filter cooling

modulate the beam coherently (in position for a transverse kicker, in energy and density for a longitudinal one). The modulation is smoothed by mixing, but some of it will always remain at the pick-up, closing the feedback loop.

The beam response is a well-known effect from the theory of instabilities in accelerator rings. For cooling purposes, because the exciting and detecting points are separated in space[5,9], the treatment is slightly different. This is not the place to discuss the details; it may, however, be said that the response as a function of frequency can be calculated if the particle distribution versus revolution frequency is given, as well as some of the ring parameters.

It is found that for separated Schottky bands and with negligible thermal noise the optimum gain for cooling corresponds to an open-loop gain with an absolute value of unity and that the phase angle of the amplifier chain response must be opposite to the phase of the beam response*. As a result of this, it turns out that in the centre of the distribution the optimum loop gain becomes - 1 for transverse cooling. The coherent feedback will then halve the amplitude of the Schottky signals as soon as the system is switched on. This is a convenient way of adjusting the gain; the correct phase may be checked by interrupting the loop somewhere and measuring its complete response with a network analyser [10].

8. Longitudinal cooling

So far, I have mainly discussed transverse cooling, i.e. reducing the betatron oscillations. Longitudinal cooling reduces the energy spread and increases the longitudinal density. This process, as it turns out, is most important for accumulating antiprotons.

One method of longitudinal cooling (sometimes called “Palmer cooling”[11]) is very similar to the one of Fig. 2. Again, we use a differential pick-up, now placed at a point where the dispersion is high, so that the particle position depends strongly on its momentum. The kicker must now give longitudinal kicks.

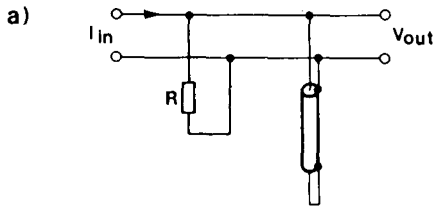


Fig. 8a. Simple transmission-line filter

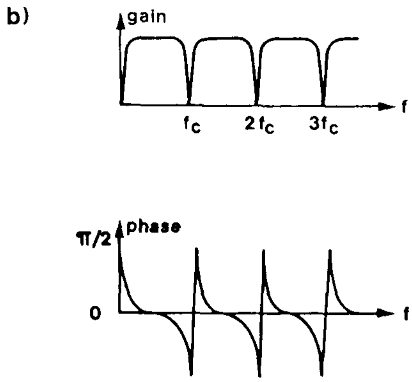


Fig. *8b.* Amplitude and phase response vs. frequency.

A different method is to use a sum pick-up (Fig. 7) and to discriminate between particles of different energy by inserting a filter into the system ("Thorndahl method"[12]). This works because the Schottky frequencies of particles with different energy are different; the filter must cause a phase change of 180° in the middle of each band, so that particles from both sides will be pushed towards the centre. Such a filter may be made by using transmission lines whose properties vary periodically with frequency. The simple filter of Fig. 8a may serve as an example. The line, shorted at the far end, behaves as a short-circuit at all resonant frequencies, which may be made to coincide with the centres of the Schottky bands. Just above these frequencies the line behaves as an inductance, just below as a capacitance; thus, the phase jump of 180" is achieved (Fig. 8b). For relativistic particles, the length of the line must be equal to half the ring's circumference. More complicated filters, using several lines and/or active feedback circuits may sometimes be useful[10].

The advantage of the filter method, especially for low-intensity beams, is that the attenuation at the central frequencies is now obtained after the preamplifier, instead of before it as with a difference pick-up. The signal-to-noise ratio is therefore much better. Also, at frequencies below about 500 MHz where ferrites may be used, sum pick-ups may be made much shorter than differential ones, so that more may fit into the same space. This again gives a better signal-to-noise ratio. Of course, for filter cooling to be practical, the Schottky bands must be separated (bad mixing).

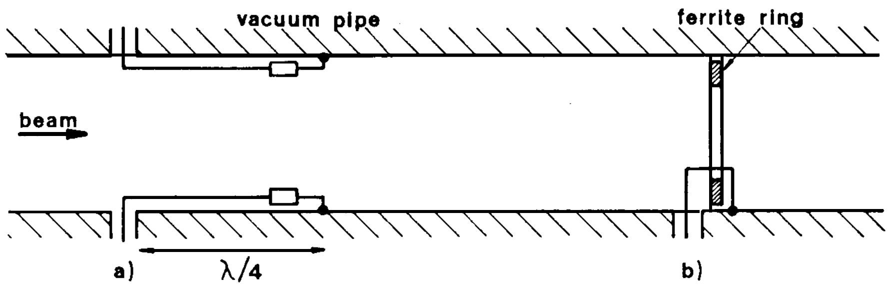


Fig. 9. Loop-type and ferrite ring-type pick-ups (or kickers). Note that for loop-type kickers the beam direction should be inverted.

9. Pick-ups and kickers

Cooling systems often have an octave bandwidth, with the highest frequency equal to twice the lowest one. Pick-ups with a reasonably flat response may consist of coupling loops that are a quarter wavelength long in the middle of the band (Fig. 9a). At the far end, a matching resistor equal to the characteristic impedance prevents reflections (or, seen in the frequency domain, ensures a correct phase relationship between beam and signal). Two loops at either side of the beam may be connected in common or differential mode for use as a sum or differential pick-up. The same structure may function as pick-up or kicker. Sum pick-ups or kickers may also consist of a ferrite frame with one or more coupling loops around it (Fig. 9b).

At high frequencies (typically > 1 GHz), slot-type pick-ups or kickers[13] become interesting (Fig. 10). The field from the particles couples to the transmission line behind the slots. If the latter are shorter than h/2, the coupling is weak and the contributions from each slot may all be added together, provided the velocity along the line is equal to the particle velocity.

The signal-to-noise ratio at the pick-ups may be improved by using many of these elements and adding their output power in matched combiner circuits. A

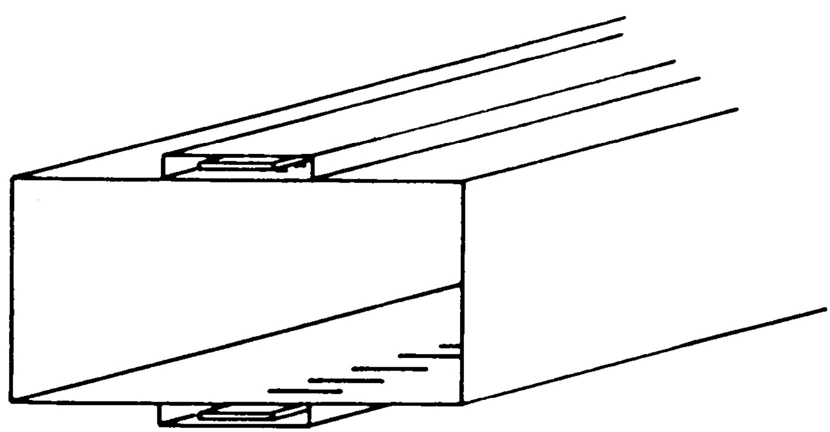


Fig. IO. Slot-type pick-up or kicker. One end of the transmission line is terminated with its own characteristic impedance.

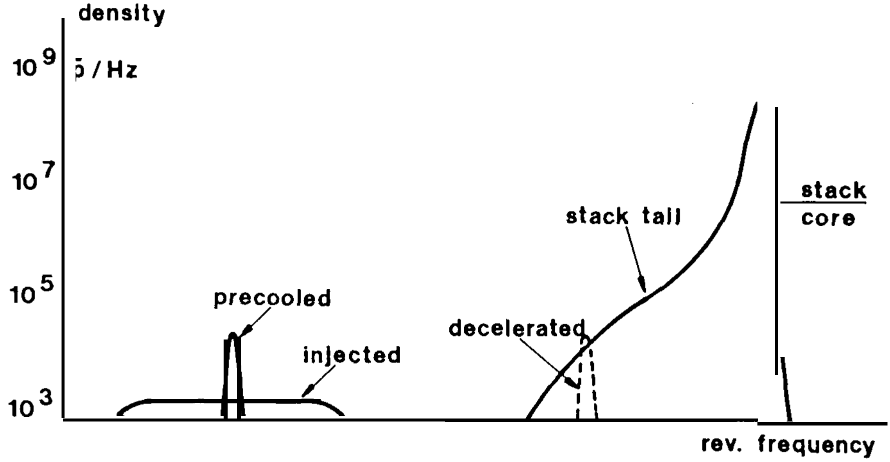


Fig. 11. Density distribution vs. revolution frequency in the Antiproton Accumulator. On the right, the stack; on the left, the newly injectcd batch, before and after precooling.

further improvement may be obtained by cryogenic cooling of the matching resistors and/or the preamplifiers.

Using many kickers reduces the total power required. The available power is sometimes a limitation to the cooling rate that may be obtained.

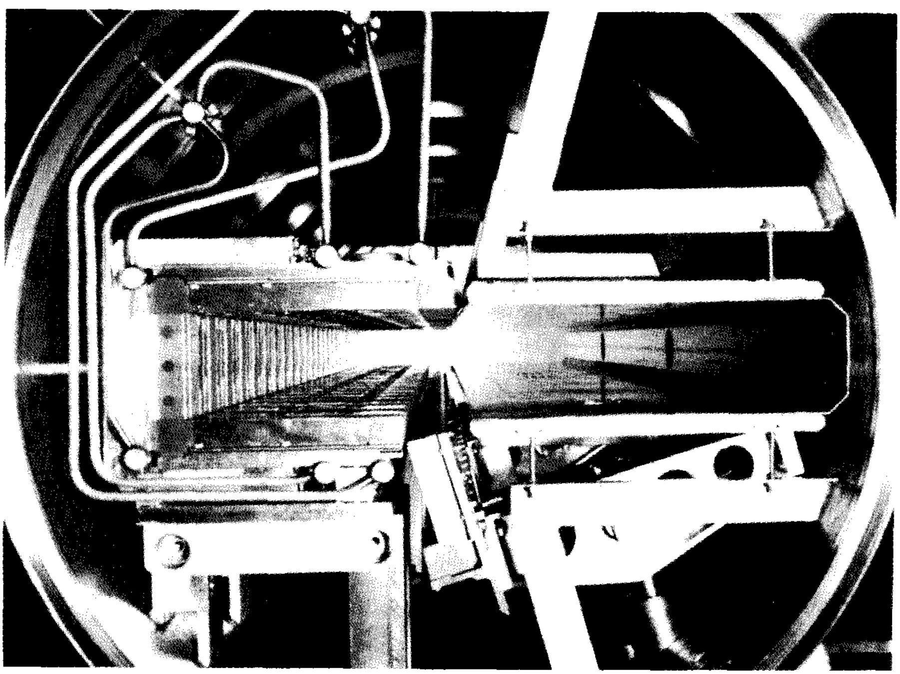


Fig. 12. Inside of a vacuum tank with precooling kickers at the left and space for the stack at the right. The fcrrite frames of the kickers arc open in the centre of the picture; they can be closed by the ferrite slabs mounted on the shutter that rotates around a pivot at the far right. Water tubes for cooling the ferrite may be seen.

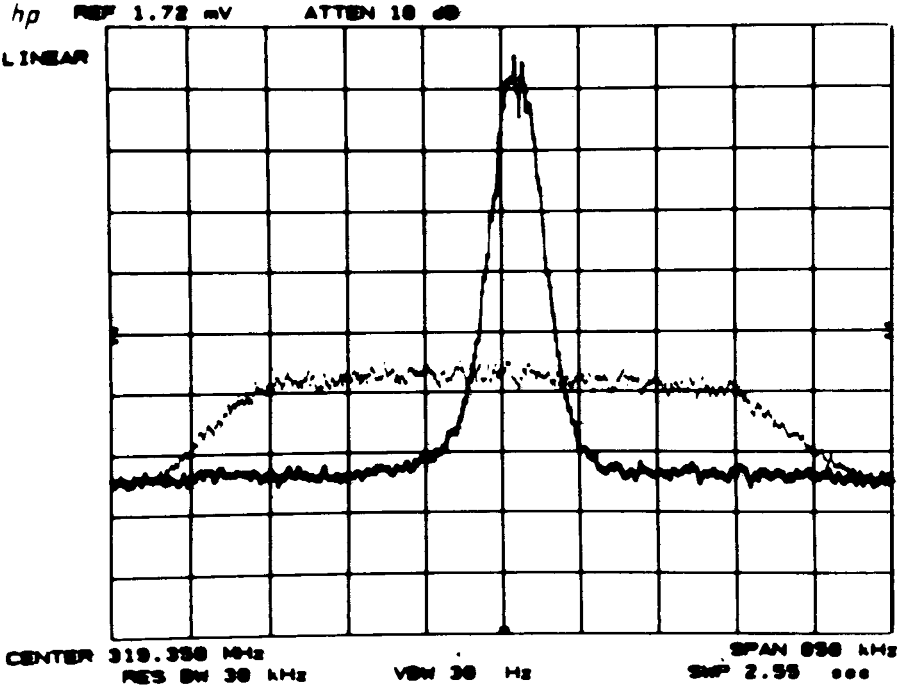


Fig. 13. Precooling 6 x 10^6 $\bar{p}$'s in 2 seconds. Longitudinal Schottky band at the 170th harmonic (314 MHz) before and after cooling.

10. *Accumulation of antiprotons; stochastic stacking*

It is now possible to explain how the antiproton accumulator works. It should, however, be made clear first that stochastic cooling is not the only method available for this purpose. In fact, already in 1966, Budker[14] proposed a pp collider scheme where the cooling was to be done by his so-called electron cooling method. A cold electron beam superimposed on the $\bar{p}$ beam cools it by electromagnetic interaction (scattering). We originally also planned to use this idea; it turns out, however, that it needs particles with low energy to work well with large-emittance beams. An additional ring to decelerate the antiprotons would then have been needed. The simpler stochastic method, using a single ring at fixed field was preferred.

In Fig. 11 we see how the particle density depends on revolution frequency (or energy, or position of the central orbit; the horizontal axis could represent any of these). On the right, the so-called stack, i.e. the particles that have already been accumulated. On the left, the low-density beam that is injected every 2.4 seconds. The latter is separated in position from the stack in those regions of the circumference where the dispersion of the lattice is large. In such a place the injection kicker can therefore inject these particles without kicking the stack. Also, the pick-ups and kickers used for the first cooling operation (longitudinal precooling) are placed here so that they do not see the stack. They consist, in fact, of ferrite frames surrounding the injected beam (Fig. 12). The pick-ups are therefore sum pick-ups (200 in total, each 25 mm long in beam direction) and the Thorndahl type of cooling, with a filter, is used[15]. Figure 13 shows how the distribution is reduced in width by an order of magnitude within 2 seconds. The number of antiprotons involved is about 6×10^6, the band used is 150-500 MHz.

After this precooling, one leg of the ferrite frames is moved downwards by a fast actuator mechanism[16] so that the precooled beam can be bunched by RF and decelerated towards the low-frequency tail of the stack (Fig. 11). The whole process, including the upward movement of the “shutter” to restore the pick-ups and kickers, takes 400 ms. The RF is then slowly reduced][17] so that the particles are debunched and deposited in the stack tail.

They must be removed from this place within the next 2.4 seconds because Liouville’s theorem prevents the RF system from depositing the next batch at the same place without simultaneously removing what was there before. A further longitudinal cooling system, using the 250-500 MHz band, therefore pushes these particles towards higher revolution frequencies, up against the density gradient[18].

This so-called stack tail system should have a gain that depends on energy (or revolution frequency). In fact, the density gradient increases strongly towards the stack core (note the logarithmic scale), and the gain for optimum cooling should vary inversely with this. We achieve this by using as pick-ups small quarter-wave coupling loops, positioned underneath and above the tail region, in such a place that they are sensitive to the extreme tail, but much less to the far-away dense core. This results in a bad signal-to-noise ratio for the region nearer to the core. Therefore, two sets of pick-ups are used, each at a different radial position and each with its own preamplifier and gain adjustment. With this set-up we obtain fast cooling at the stack edge where the particles are deposited, and slow cooling at the dense core, where we can afford it because the particles remain there for hours.

A problem is that the tail systems must be quite powerful to remove the particles fast enough. As a result, their kickers will also disturb the slowly-cooled stack core (the Schottky signals do not overlap with the core frequencies, but the thermal noise does). The problem exists because the kickers must be at a point where the dispersion is zero to prevent them from exciting horizontal betatron oscillations. They therefore kick all particles (tail or core) equally.

A solution is found by using transmission-line filters as described above to suppress the core frequencies in the tail cooling systems. These filters also rotate the phase near the core region in an undesirable way; this does not matter, however, because the cooling of the core is done by a third system of larger bandwidth (l-2 GHz).

While the particles move towards the core, they are also cooled horizontally and vertically, first by tail cooling systems, then by l-2 GHz core systems. The layout of the various cooling circuits is shown in Fig. 14. In the general view of Fig. 15, some of the transmission lines transporting the signals for the pick-ups to the kickers may be seen.

When the stack contains a sufficient number of antiprotons (typically 2×10^{11}), a fraction of these (~ 30 %) is transferred to the PS and from there to

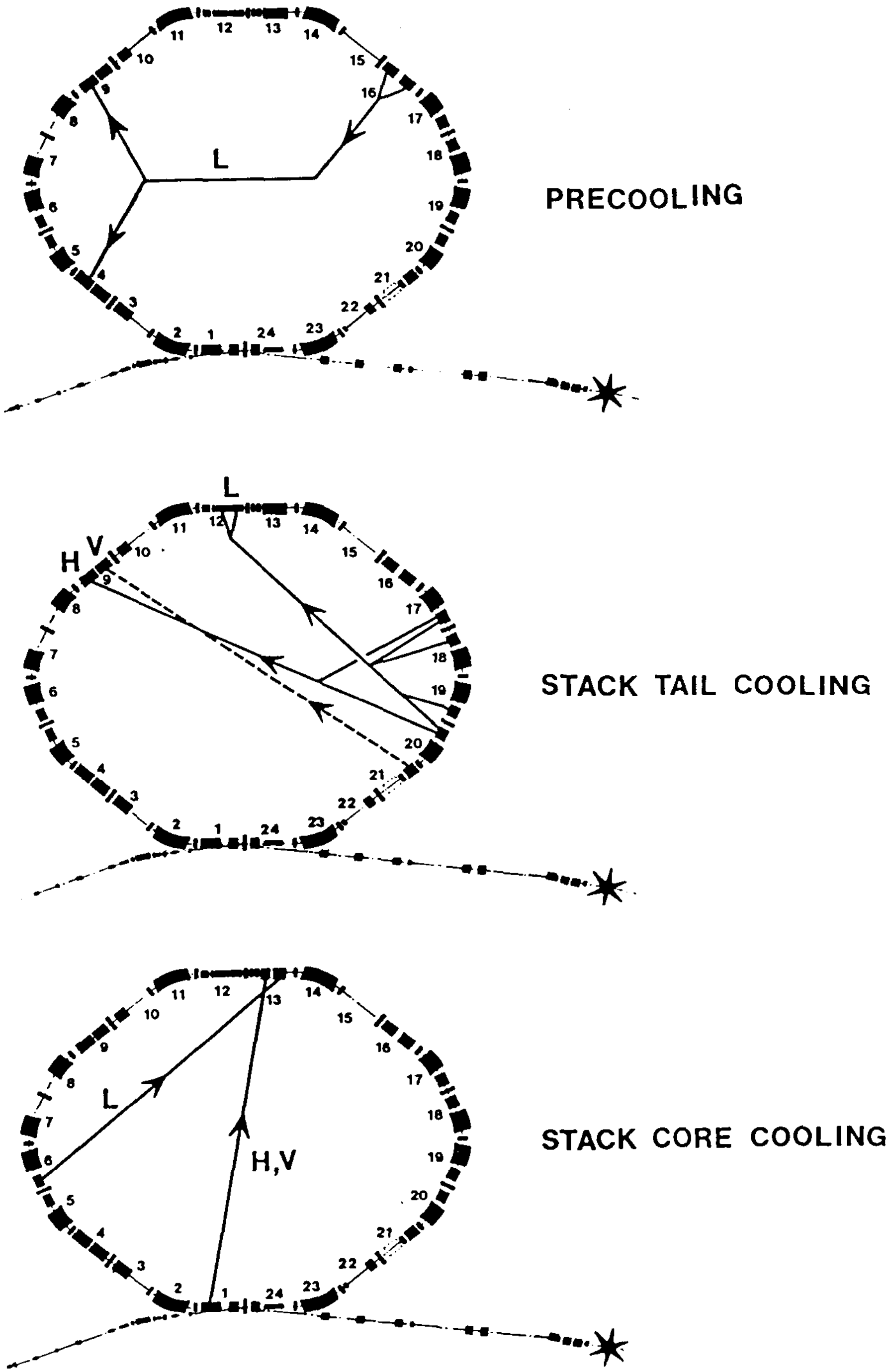


Fig. 14. Plan of the AA ring with its 7 cooling systems. L = longitudinal, V = vertical, H = horizontal.

the SPS machine. This is done by bunching a part of the stack, of a width that may be adjusted by properly choosing the RF bucket area[19]. These are accelerated until they are on the same orbit where normally particles are injected. They can then be extracted without disturbing the remaining stack. This process is repeated (at present three times); each time one RF bucket of the SPS is filled. The remaining p's form the beginning of the next stack.

11. *Design of longitudinal cooling systems; Fokker-Planck equation*

The main difference between transverse and longitudinal cooling systems is that the latter will change the longitudinal distribution on which the incoherent (heating) term depends, as well as effects such as the beam feedback. This complicates the theory; still, everything can be calculated if all parameters are given.

It is convenient to define the flux ϕ, i.e. the number of particles passing a certain energy (or frequency) value per unit time. It may be shown[5] that

$$\phi = F\Psi - D\delta\Psi/\delta f_0, \tag{5}$$

where Ψ is the density dN/df_0 while F and D are slowly varying constants, depending on various system parameters as well as on the particle distribution. The first term represents the coherent cooling, the second one the incoherent (diffusion) effect that has the effect of pushing the particles down the gradient under the influence of perturbing noise.

By using the continuity equation

$$\delta\Psi/\delta t + \delta\phi/\delta f_0 = 0,$$

expressing that no particles are lost, we find the Fokker-Planck-type equation

$$\frac{\delta\Psi}{\delta t} = -\frac{\delta}{\delta f_0}(F\Psi) + \frac{\delta}{\delta f_0}\left(D\frac{\delta\Psi}{\delta f_0}\right) \tag{6}$$

that allows us to compute the evolution of the density versus revolution frequency f_0 and time given the initial distribution. The particles deposited at the edge are introduced as a given flux at that point.

The constants F and D depend on many system parameters (pick-up and kicker characteristics, amplifier gain, filter response, beam distribution, etc.). Their value is found through summing the contributions of all Schottky bands. Analytic solutions of (6) do not exist in practice and a complicated numerical treatment is indicated.

Such calculations resulted in the design of the antiproton stacking system. At the time this was done, tests in a small experimental ring (ICE) had confirmed the cooling in all planes at time scales of the order of 10 seconds. However, it was not possible to check the stacking system (increasing the density by four orders of magnitude) in any way, and it may be argued that we took a certain risk by starting the project without being able to verify this aspect. Fortunately,

Fig. 15. View of the Antiproton Accumulator beforc it was covered by concrete slabs. The silvered material around the vacuum tanks is insulation, needed because everything may be heated to 300” to obtain ultra-high vacuum. The transmission lines crossing the ring and carrying the cooling signals may be seen.

everything behaved according to theory and although the number of p’s injected is smaller than was hoped for by a factor 3.5, the cooling works largely as expected.

12. Other applications of stochastic cooling; future developments

At present. stochastic cooling is used at CERN in the $\bar{p}$ accumulator and in the low energy ring (LEAR) where the p’s may be stored after deceleration in the PS. Before the intersecting storage rings (ISR) were closed down last year, they also used the antiprotons and contained cooling equipment.

In the SPS where the high-energy collisions take place, cooling would be attractive because it would improve the beam lifetime and might decrease its cross-section. However, a difficulty is formed by the fact that the beam is bunched in this machine; the bunches are narrow (3 x 4 ns). In fact, owing to the bunching each Schottky band is split up into narrow, dense satellite bands and the signals from different bands are correlated”. Nevertheless, a scheme is being considered that might improve the lifetime to a certain extent”.

In the United States, a $\bar{p}$ accumulator complex similar to the CERN one and also using stochastic cooling is being constructed”. This machine is expected to have a stacking rate an order of magnitude higher than the CERN one because it uses a higher primary energy to produce the antiprotons and higher frequencies to cool them. In the meantime, we are building a second ring at CERN, surrounding the present accumulator (Fig. 16), with a similar performance. It will have stronger focusing, so increasing both transverse acceptances by at least a factor 2, and the longitudinal one by a factor 4. The increased focusing strengths will diminish the mixing; consequently, higher frequencies (up to 4 GHz) will be used for cooling. The present AA will be used to contain the stack and its cooling systems will also be upgraded.

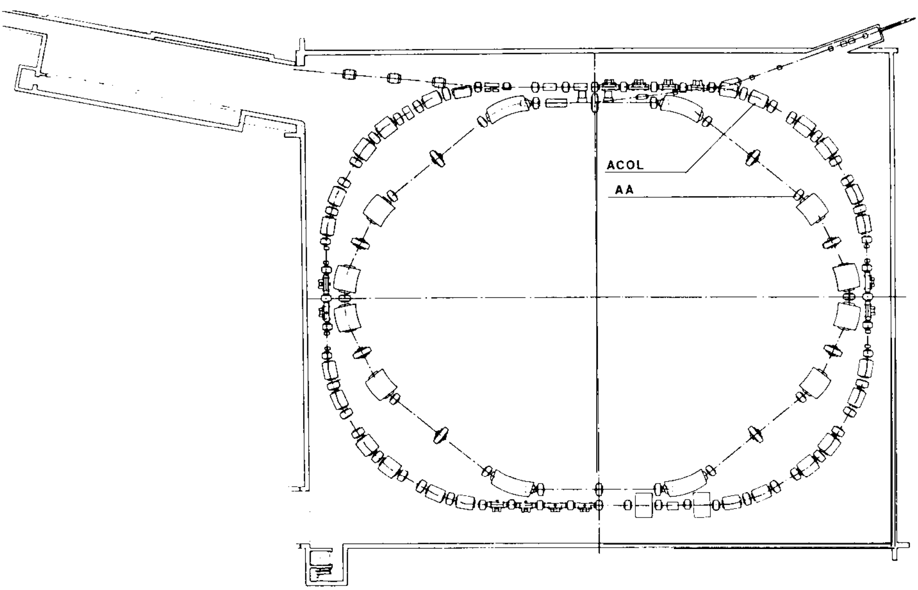


Fig. 16. The new ACOL ring (under construction) around the AA. This ring will increase the stacking rate by an order of magnitude. The stack will still be kept in the AA ring.

ACKNOWLEDGEMENTS

The development of the stochastic cooling. theory owes much to H. G. Hereward, D. Möhl, F. Sacherer, and L. Thorndahl. The latter also made important contributions to the construction of most cooling systems at CERN and it is doubtful if the accumulator would have been feasible without his invention of the filter method. It is also a pleasure to acknowledge the invaluable contributions of G. Carron (hardware), L. Faltin (slot pick-ups) and C. Taylor (l-2 GHz systems).

During the construction of the Antiproton Accumulator, R. Billinge was joint project leader with myself and it is mostly because of his contributions to the design and his able management that the machine was ready in a record time (2 years) and worked so well.

REFERENCES

1. L. Szilard, Gbcr die Entropieverminderung in &cm thermodynamischen System bei Eingriffin intclligenter Wesen, Zcitsch. f. Physik 53 (1929) 840.
2. S. van dcr Meer, Stochastic damping of betatron oscillations in the ISR, CERN/ISR-PO/ 72-31 (1972).
3. D. Möhl, Stochastic cooling for beginners, Proceedings of the CERN Accelerator School on Antiprotons for Colliding Beam Facilities, CERN 84- 15 (1984).
4. F. Sacherer, Stochastic cooling theory, CERN/ISR-TH/78- 11 (1978).
5. D. Möhl, G. Petrucci, L. Thorndahl, S. van der Merr, Physics and technique of stochastic cooling, Phys. Reports 58 (1980) 73.
6. W. Schottky, Ann. Physik 57 (1918) 541.
7. H. G. Hereward, The elementary theory of Landau damping, CERN 65-20 (1965).
8. S. van der Meer, Optimum gain and phase for stochastic cooling systems, CERN/PS-AA/83-48 (1983).
9. S. van der Meer. A different formulation of the longitudinal and transvcrse beam response, CERN/PS-AA/80-4 (1980).
10. S. van der Meer, Stochastic cooling in the Antiproton Accumulator, IEEE Trans. Nucl. Sci. NS28 (1981) 1994.
11. R. B. Palmer, BNL, Private communication (1975).
12. G. Carron, L. Thorndahl, Stochastic cooling of momentum spread by filter techniques, CERN/ISR-RF/78-12 (1978).
13. L. Faltin, Slot-type pick-up and kicker for stochastic beam cooling, Nucl. Instrum. Methods 148 (1978) 449.
14. G. I. Budker, Proc. Int. Symp. on Electron and Positron Storage Rings, Saclay, 1966, p. 11 -l-1; Atom. Energ. 22 (1967) 346.
    G. I. Budkcr et al., Experimental studies of electron cooling, Part. Acc. 7 (1976) 197.
15. S. van der Meer, Precooking in the Antiproton Accumulator, CERN/PS-AA/78-26 (1978).
16. D. C. Fiander, S. Milner, P. Pearce, A. Poncet, The Antiproton Accumulator shutters: design, technology and performance, CERN/PS/84-23 (1984).
17. R. Johnson, S. van dcr Meer, F. Pcdersen, G. Shering, Computer control of RF manipulations in the CERN Antiproton Accumulator, IEEE Trans. Nucl. Sci. NS-30 (1983) 2290.
18. S. van der Meer, Stochastic stacking in the Antiproton Accumulator, CERN/PS-AA/78-22 (1978).
19. R. Johnson, S. van dcr Meer, F. Pedersen, Measuring and manipulating an accumulated stack of antiprotons in the CERN Antiproton Accumulator, IEEE Trans. Nucl. Sci. NS-30 (1983) 2123.
20. H. Herr, D. Möhl, Bunched bram stochastic cooling, CERN/EP/Note 79-34 (1979).
21. D. Boussard, S. Chattopadhyay, G. Dôme, T. Linnecar, Feasibility study of stochastic cooling of bunches in the SPS, CERN/SPS/84-4 (1984).
22. Design Report Tevatron I project, Fermi National Accelerator Laboratory, Batavia, 111. (1983).

Physics 1985

KLAUS VON KLITZING

for the discovery of the quantized Hall effect

THE NOBEL PRIZE FOR PHYSICS

Speech by Professor STIG LUNDQVIST of the Royal Academy of Sciences.
Translation from the Swedish text

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen,

This year's Nobel Prize for Physics has been awarded to Professor Klaus von Klitzing for the discovery of the quantized Hall effect.

This discovery is an example of these unexpected and surprising discoveries that now and then take place and which make research in the sciences so exciting. The Nobel Prize is sometimes an award given to large projects, where one has shown great leadership and where one with ingenuity combined with large facilities and material resources has experimentally verified the correctness of theoretical models and their predictions. Or, one has succeeded through creation of new theoretical concepts and methods to develop theories for fundamental problems in physics that resisted all theoretical attempts over a long period of time. However, now and then things happen in physics that no one can anticipate. Someone discovers a new phenomon or a new fundamental relation in areas of physics where no one expects anything exciting to happen.

This was exactly what happened when Klaus von Klitzing in February 1980 was working on the Hall effect at the Hochfelt-Magnet-Labor in Grenoble. He discovered from his experimental data that a relation which had been assumed to hold only approximately seemed to hold with an exceptionally high accuracy and in this way the discovery of the quantized Hall effect was made.

The discovery by von Klitzing has to do with the relation between electric and magnetic forces in nature and has a long history. Let us go back to 1820, when the Danish physicist H. C. Ørsted found that an electric current in a wire influenced a compass needle and made it change its direction. He discovered this phenomenon in a class with his students. No one had seen a relation between electric and magnetic forces before. More than 50 years later a young American physicist, E. H. Hall, speculated that the magnetic force might influence the charge carriers in a metallic wire placed in a magnetic field and give rise to an electric voltage across the wire. He was able to show that when sending an electric current through a strip of gold there was a small voltage across the wire in a direction perpendicular both to the current and the magnetic field. That was the discovery of the Hall effect.

The Hall effect is now a standard method frequently used to study semiconductor materials of technical importance, and the effect is described in all textbooks in solid state physics. The experiment is in principle very simple and requires only a magnetic field plus instruments to measure current and voltage. If one varies the magnetic field, the current and voltage will change in a completely regular way and no surprising effects are expected to happen.

von Klitzing studied the Hall effect under quite extreme conditions. He used an extremely high magnetic field and cooled his samples to just a couple of degrees above the absolute zero point of temperature. Instead of the regular change one would expect, he found some very characteristic steps with plateaus in the conductivity. The values at these plateaus can with extremely high accuracy be expressed as an integer times a simple expression that just depends of two fundamental constants: the electric elementary charge and Planck's constant which appear everywhere in quantum physics.

The result represents a quantization of the Hall effect-a completely unexpected effect. The accuracy in his results was about one part in ten million, which would correspond to measuring the distance between Stockholm and von Klitzing's home station Stuttgart with an accuracy of a few centimeters.

The discovery of the quantized Hall effect is a beautiful example of the close interrelation between the highly advanced technology in the semiconductor industry and fundamental research in physics. The samples used by von Klitzing were relined versions of a kind of transistor we have in our radios. His samples, however, had to satisfy extremely high standards of perfection and could only be made by using a highly advanced technique and refined technology.

The quantized Hall effect can only be observed in a two-dimensional electron system. Two-dimensional electron systems do not occur in nature. However, the development in semiconductor technology has made possible the realization of a two-dimensional electron system. In the kind of transistor that von Klitzing used, some of the electrons are bound to the interface between two parts of the transistor. At sufficiently low temperature the electrons can move only along the interface and one has effectively a two-dimensional electron system.

von Klitzing's discovery of the quantized Hall effect attracted immediately an enormous interest. Because of the extremely high accuracy the effect can be used to define an international standard for electric resistance. The metrological possibilities are of great importance and have been subject to detailed studies at many laboratories all over the world.

The quantized Hall effect is one of the few examples, where quantum effects can be studied in ordinary macroscopic measurements. The underlying detailed physical mechanisms are not yet fully understood. Later experiments have revealed completely new and unexpected properties and the study of two-dimensional systems is now one of the most challenging areas of research in physics.

Professor von Klitzing,

On behalf of the Royal Swedish Academy of Sciences I wish to convey our warmest congratulations and ask you to receive your prize from the hands of His Majesty the King.

KLAUS v. KLITZING

Born 28th June 1943 in Schroda (Posen), German nationality.
February 1962:
Abitur in Quakenbrück
April 1962 to March 1969:
Technical University Braunschweig Diploma in Physics.
Title of diploma work: “Lifetime Measurements on InSb” (Prof. F. R. Kepler)
May 1969 to Nov. 1980:
University Würzburg (Prof. Dr. G. Landwehr)
Thesis work about: “Galvanomagnetic Properties of Tellurium in Strong Magnetic Fields” (Ph.D. in 1972).
Habilitation in 1978.
The most important publication related to the Nobel Prize appeared in: Phys. Rev. Letters 45, 494 (1980).
Research work at the Clarendon Laboratory, Oxford (1975 to 1976) and High Magnetic Field Laboratory, Grenoble (1979 to 1980)
Nov. 1980 to Dec. 1984:
Professor at the Technical University, München
Since January 1985:
Director at the Max-Planck-Institut für Festkörperforschung, Stuttgart.

THE QUANTIZED HALL EFFECT

Nobel lecture, December 9, 1985

by

KLAUS von KLITZING,

Max-Planck-Institut für Festkörperforschung, D-7000 Stuttgart 80

1. *Introduction*

Semiconductor research and the Nobel Prize in physics seem to be contradictory since one may come to the conclusion that such a complicated system like a semiconuctor is not useful for very fundamental discoveries. Indeed, most of the experimental data in solid state physics are analyzed on the basis of simplified theories, and very **often the properties of a semiconductor device is** described by empirical formulas since the microscopic **details are too complicated.** Up to 1980 nobody expected that there exists an effect like the Quantized Hall Effect, which depends exclusively on fundamental constants and is not affected by irregularities in the semiconductor like impurities or interface effects.

The discovery of the Quantized Hall Effect (QHE) was the result of systematic measurements on silicon field effect transistors-the most important device in microelectronics. Such devices are not only important for applications but also for basic research. The pioneering work by Fowler, Fang, Howard and Stiles [l] has shown that new quantum phenomena become visible if the electrons of a conductor are confined within a typical length of 10 nm. Their discoveries opened the field of two-dimensional electron systems which since 1975 is the subject of a conference series [2]. It has been demonstrated that this field is important for the description of nearly all optical and electrical properties of microelectronic devices. A two-dimensional electron gas is absolutely necessary for the observation of the Quantized Hall Effect, **and the realization** and properties of such a system will be discussed in section 2. In addition to the quantum phenomena connected with the confinement of electrons within a two-dimensional layer, another quantization-the Landau quantization of the electron motion in a strong magnetic field-is essential for the interpretation of the Quantized Hall Effect (section 3). Some experimental results will be summarized in section 4 and the application of the QHE in metrology is the subject of section 5.

2 Two-Dimensional Electron Gas

The fundamental properties of the QHE are a consequence of the fact that the energy spectrum of the electronic system used for the experiments is a *discrete* **energy spectrum. Normally, the energy E of mobile electrons in a** semiconductor is quasicontinuous and can be compared with the kinetic energy of free electrons with wave vector k but with an effective mass m*

$$E = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2 + k_z^2) \tag{1}$$

If the energy for the motion in one direction (usually z-direction) is fixed, one obtains a quasi-two-dimensional electron gas (2DEG), and a strong magnetic field perpendicular to the two-dimensional plane will lead-as discussed later- to a fully quantized energy spectrum which is necessary for the observation of the QHE.

A two-dimensional electron gas can be realized at the surface of a semiconductor like silicon or gallium arsenide where the surface is usually in contact with a material which acts as an insulator (SiO2 for silicon field effect transistors and, e.g. $Al_xGa_{1-x}As$ for heterostructures). Typical cross sections of such devices are shown in Fig 1. Electrons are confined close to the surface of the semiconductor by an electrostatic field F_z normal to the interface, originating from positive charges (see Fig. 1) which causes a drop in the electron potential towards the surface.

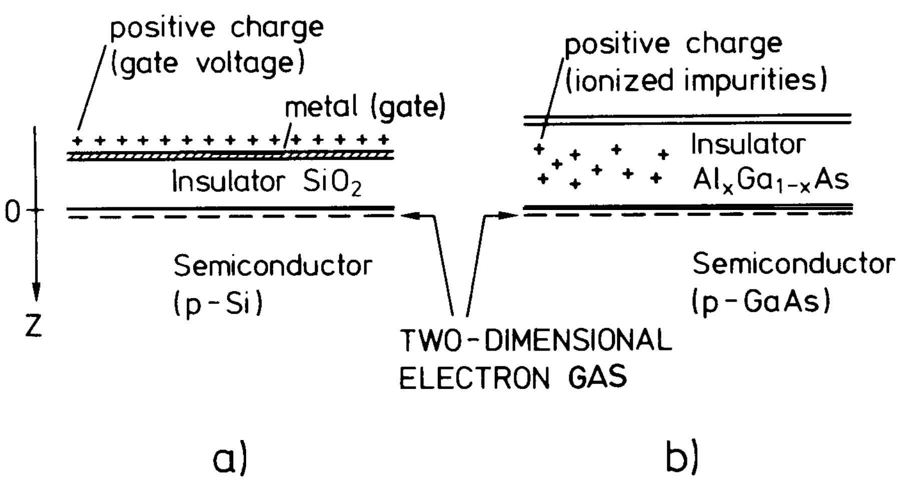


Fig. I. A two-dimensional electron gas (2DEG) can be formed at the semiconductor surface if the electrons are fixed close to the surface by an external electric field. Silicon MOSFETs (a) and GaAs-$Al_xGa_{1-x}As$ heterostructures (b) are typical structures used for the realization of a 2DEG.

If the width of this potential well is small compared to the de Broglie wavelength of the electrons, the energy of the carriers are grouped in so-called electric subbands Ei corresponding to quantized levels for the motion in z-direction, the direction normal to the surface. In lowest approximation, the electronic subbands can be estimated by calculating the energy eigenvalues of an electron in a triangular potential with an infinite barrier at the surface (z=0) and a constant electric field F_s for $z \geq 0$, which keeps the electrons close to the surface. The result of such calculations can be approximated by the equation

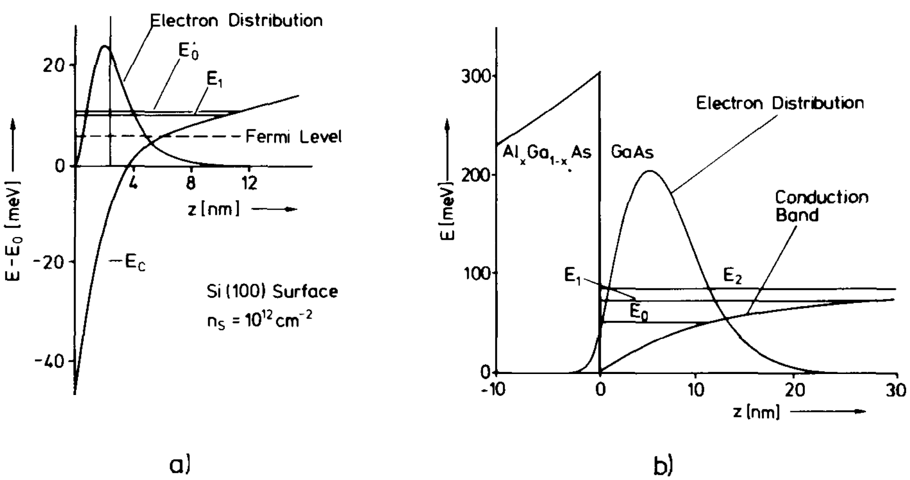
$$E_j = \left(\frac{\hbar^2}{2m^*}\right)^{1/3} \cdot \left(\frac{3}{2}\pi e F_s\right)^{2/3} \cdot \left(j + \frac{3}{4}\right)^{2/3} \tag{2}$$

$j = 0, 1, 2 \ldots$

In some materials, like silicon, different effective masses m* and $m^{*'}$ may be present which leads to different series Ej and Ej.

Equation (2) must be incorrect if the energy levels Ej are occupied with electrons, since the electric field F_s will be screened by the electronic charge.

For a more quantitative calculation of the energies of the electric subbands it is necessary to solve the Schrödinger equation for the actual potential V'_z which changes with the distribution of the electrons in the inversion layer. Typical results of such calculation for both silicon MOSFETs and GaAs-heterostructures are shown in Fig. 2 [3,4]. Usually, the electron concentration of the two-dimensional system is fixed for a heterostructure (Fig. 1 b) but can be varied in a MOSFET by changing the gate voltage.



Fig, 2. Calculations of the electric subbands and the electron distribution within the surface channel of a silicon MOSFET (a) and a GaAs-$Al_xGa_{1-x}As$ heterostructure [3, 4].

Experimentally, the separation between electric subbands, which is of the order of 10 meV, can be measured by analyzing the resonance absorption of electromagnetic waves with a polarization of the electric field perpendicular to the interface [5].

At low temperatures (T<4 K) and small carrier densities for the 2DEG (Fermi energy E_F relative to the lowest electric subbands E_0 small compared with the subband separation E_1-E_0) only the lowest electric subband is occupied with electrons (electric quantum limit), which leads to a strictly two-dimensional electron gas with an energy spectrum

$$E = E_0 + \frac{\hbar^2 k_{\parallel}^2}{2m^*} \tag{3}$$

where $k_{\parallel}$ is a wavevector within the two-dimensional plane.

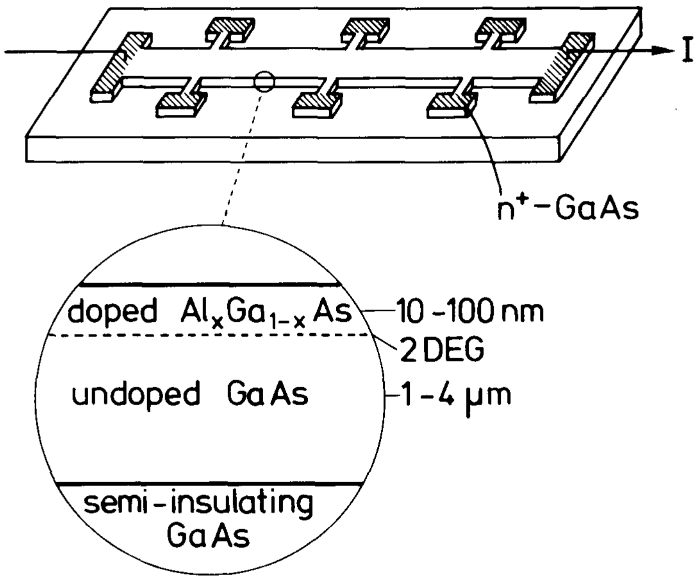


Fig. 3. Typical shape and cross-section of a GaAs-$Al_xGa_{1-x}As$ hcterostructure used for Hall effect measurements.

For electrical measurements on a 2DEG, heavily doped n^+-contacts at the semiconductor surface are used as current contacts and potential probes. The shape of a typical sample used for QHE-experiments (GaAs-heterostructure) is shown in Fig. 3. The electrical current is flowing through the surface channel, since the fully depleted $Al_xGa_{1-x}As$ acts as an insulator (the same is true for the SiO_2 of a MOSFET) and the p-type semiconductor is electrically separated from the 2DEG by a p-n junction. It should be noted that the sample shown in Fig. 3 is basically identical with new devices which may be important for the next computer generation [6]. Measurements related to the Quantized Hall Effect which include an analysis and characterization of the 2DEG are therefore important for the development of devices, too.

3. Quantum Transport of a 2DEG in Strong Magnetic Fields

A strong magnetic field B with a component B_z, normal to the interface causes the electrons in the two-dimensional layer to move in cyclotron orbits parallel to the surface. As a consequence of the orbital quantization the energy levels of the 2DEG can be written schematically in the form

$$E_n = E_o + (n + 1/2)\hbar\omega_c + s \cdot g \cdot \mu_B \cdot B \qquad (4)$$
$$n = 0,1,2,...$$

with the cyclotron energy $\hbar\omega_c = \hbar eB/m^*$, the spin quantum numbers $s = \pm 1/2$ the Landé factor g and the Bohr magneton μ_B

The wave function of a 2DEG in a strong magnetic field may be written in a form where the y-coordinate y_0 of the center of the cyclotron orbit is a good quantum number [7].

$$\psi = e^{ikx}\Phi_n(y-y_o) \qquad (5)$$

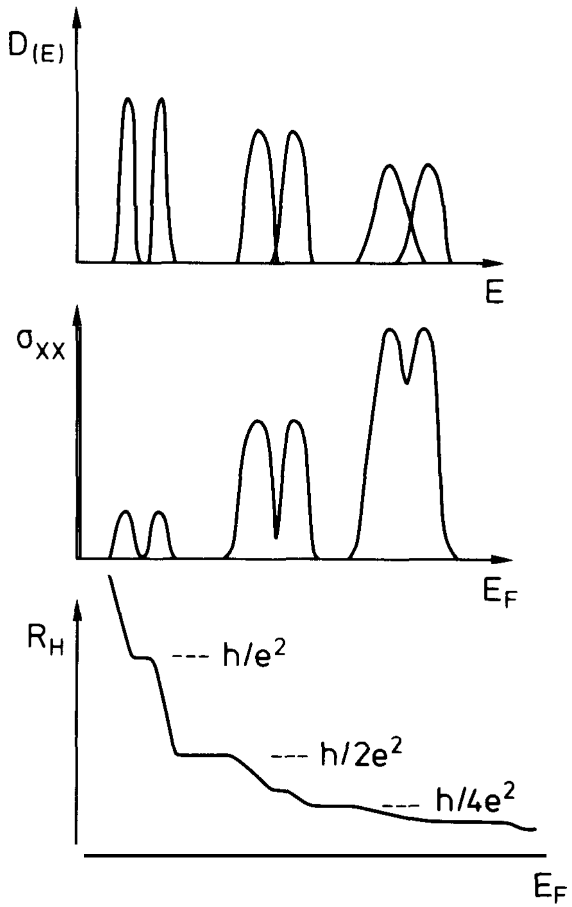


Fig. 4. Sketch for the energy dependence of the density of states (a), conductivity σ_{xx} (b), and Hall resistance R_H(c) at a fixed magnetic field.

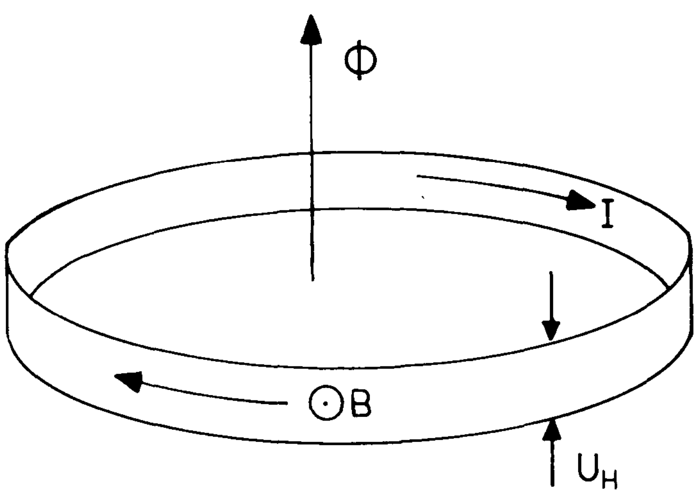


Fig. 5. Model of a two-dimensional metallic loop used for the derivation of the quantized Hall resistance.

where Φ_n is the solution of the harmonic-oscillator equation

$$\frac{1}{2m^*}\left[p_y^2 + (eB)^2y^2\right] \phi_n = (n + \tfrac{1}{2})\,\hbar\omega_c\,\phi_n \tag{6}$$

and y_0 is related to k by

$$y_0 = \hbar k/eB \tag{7}$$

The degeneracy factor for each Landau level is given by the number of center coordinates y_0. within the sample. For a given device with the dimension $L_x \cdot L_y$, the center coordinates y_0 are separated by the amount

$$\Delta y_0 = \frac{\hbar}{eB}\Delta k = \frac{\hbar}{eB}\frac{2\pi}{L_x} = \frac{h}{eBL_x} \tag{8}$$

so that the degeneracy factor $N_0 = L_y/\Delta y_0$ is identical with $N_0 = L_xL_yeB/h$ the number of flux quanta within the sample. The degeneracy factor per unit area is therefore:

$$N = \frac{N_0}{L_xL_y} = \frac{eB}{h} \tag{9}$$

It should be noted that this degeneracy factor for each Landau level is independent of semiconductor parameters like effective mass.

In a more general way one can show [8] that the commutator for the center coordinates of the cyclotron orbit $[x_0,y_0] = i\hbar/eB$ is finite, which is equivalent to the result that each state occupies in real space the area $F_0 = h/eB$ corresponding to the area of a flux quantum.

The classical expression for the Hall voltage U_H of a 2DEG with a surface carrier density n_s is

$$U_H = \frac{B}{n_s \cdot e} \cdot I \tag{10}$$

where I is the current through the sample. A calculation of the Hall resistance $R_H = U_H/I$ under the condition that i energy levels are fully occupied ($n_s = iN$), leads to the expression for the quantized Hall resistance

$$R_H = \frac{B}{iN \cdot e} = \frac{h}{ie^2} \tag{11}$$

$$i = 1, 2, 3$$

A quantized Hall resistance is always expected if the carrier density n_s, and the magnetic field B are adjusted in such a way that the filling factor i of the energy Levels (Eq. 4)

$$i = \frac{n_s}{eB/h} \tag{12}$$

is an integer.

Under this condition the conductivity σ_{xx} (current flow in the direction of the electric field) becomes zero since the electrons are moving like free particles exclusively perpendicular to the electric field and no diffusion (originating from scattering) in the direction of the electric field is possible. Within the self-consistent Born approximation [9] the discrete energy spectrum broadens as shown in Fig. 4a. This theory predicts that the conductivity σ_{xx} is mainly proportional to the square of the density of states at the Fermi energy E_F which leads to a vanishing conductivity σ_{xx} in the quantum Hall regime and quantized plateaus in the Hall resistance R_H (Fig. 4c).

The simple one-electron picture for the Hall effect of an ideal two-dimensional system in a strong magnetic field leads already to the correct value for the quantized Hall resistance (Eq. 11) at integer filling factors of the Landau levels. However, a microscopic interpretation of the QHE has to include the influences of the finite size of the sample, the finite temperature, the electron-electron interaction, impurities and the finite current density (including the inhomogenious current distribution within the sample) on the experimental result. Up to now, no corrections to the value h/ie^2 of the quantized Hall resistance are predicted if the conductivity σ_{xx} is zero. Experimentally, σ_{xx} is never exactly zero in the quantum Hall regime (see section 4) but becomes unmeasurably small at high magnetic fields and low temperatures. A quantitative theory of the QHE has to include an analysis of the longitudinal conductivity σ_{xx} under real experimental conditions, and a large number of publications are discussing the dependence of the conductivity on the temperature, magnetic field, current density, sample size etc. The fact that the value of the quantized Hall resistance seems to be exactly correct for $\sigma_{xx} = 0$ has led to the conclusion that the knowledge of microscopic details of the device is not necessary for a calculation of the quantized value. Consequently Laughlin [10] tried to deduce the result in a more general way from gauge invariances. He considered the situation shown in Fig. 5. A ribbon of a two-dimensional system is bent into a loop and pierced everywhere by a magnetic field B normal to its surface. A voltage drop U_H is applied between the two edges of the ring. Under the condition of vanishing conductivity σ_{xx} (no energy dissipation), energy is conserved and one can write Faraday's law ofinduction in a form which relates the current I in the loop to the adiabatic derivative of the total energy of the system E with respect to the magnetic flux Φ threading the loop

$$I = \frac{\partial E}{\partial \phi} \tag{13}$$

If the flux is varied by a flux quantum $\Phi_0 = h/e$, the wavefunction enclosing the flux must change by a phase factor 2π corresponding to a transition of a state with wavevector k into its neighbour state $k + (2\pi) / (L_x)$, where L_x is the circumference of the ring. The total change in energy corresponds to a transport of states from one edge to the other with

$$\Delta E = i \cdot e \cdot U_H \tag{14}$$

The integer i corresponds to the number of filled Landau levels if the free electron model is used, but can be in principle any positive or negative integer number.

From Eq. (13) the relation between the dissipationless Hall current and the Hall voltage can be deduced

$$I = i\cdot e\cdot U_H/\Phi_0 = i\,\frac{e}{\cdot}\cdot U_H \tag{15}$$

which leads to the quantized Hall resistance $R_H = \frac{h}{ie^2}$.

In this picture the main reason for the Hall quantization is the flux quantization h/e and the quantization of charge into elementary charges e. In analogy, the fractional quantum Hall effect, which will not be discussed in this paper, is interpreted on the basis of elementary excitations of quasiparticles with a charge $e^* = \frac{e}{3}, \frac{e}{5}, \frac{e}{7}$ etc.

The simple theory predicts that the ratio between the carrier density and the magnetic field has to be adjusted with very high precision in order to get exactly integer filling factors (Eq. 12) and therefore quantized values for the Hall resistance. Fortunately, the Hall quantization is observed not only at special magnetic field values but in a wide magnetic field range, so that an accurate fixing of the magnetic field or the carrier density for high precision measurements of the quantized resistance value is not necessary. Experimental data of such Hall plateaus are shown in the next section and it is believed that localized states are responsible for the observed stabilization of the Hall resistance at certain quantized values.

After the discovery of the QHE a large number of theoretical paper were published discussing the influence of localized states on the Hall effect [11 - 14] and these calculations demonstrate that the Hall plateaus can be explained if localized states in the tails of the Landau levels are assumed. Theoretical investigations have shown that a mobility edge exists in the tails of Landau levels separating extended states from localized states [15-18]. The mobility edges are located close to the center of a Landau level for long-range potential fluctuations. Contrary to the conclusion reached by Abrahams, et al [19] that all states of a two-dimensional system are localized, one has to assume that in a strong magnetic field at least one state of each Landau level is extended in order to observe a quantized Hall resistance. Some calculations indicate that the extended states are connected with edge states [17].

In principle, an explanation of the Hall plateaus without including localized states in the tails of the Landau levels is possible if a reservoir of states is present outside the two-dimensional system [20, 21]. Such a reservoir for electrons, which should be in equilibrium with the 2DEG, fixes the Fermi energy within the energy gap between the Landau levels if the magnetic field or the number of electrons is changed. However, this mechanism seems to be more unlikely than localization in the the tails of the Landau levels due to disorder. The following discussion assumes therefore a model with extended and localized states within one Landau level and a density of states as sketched in Fig. 6.

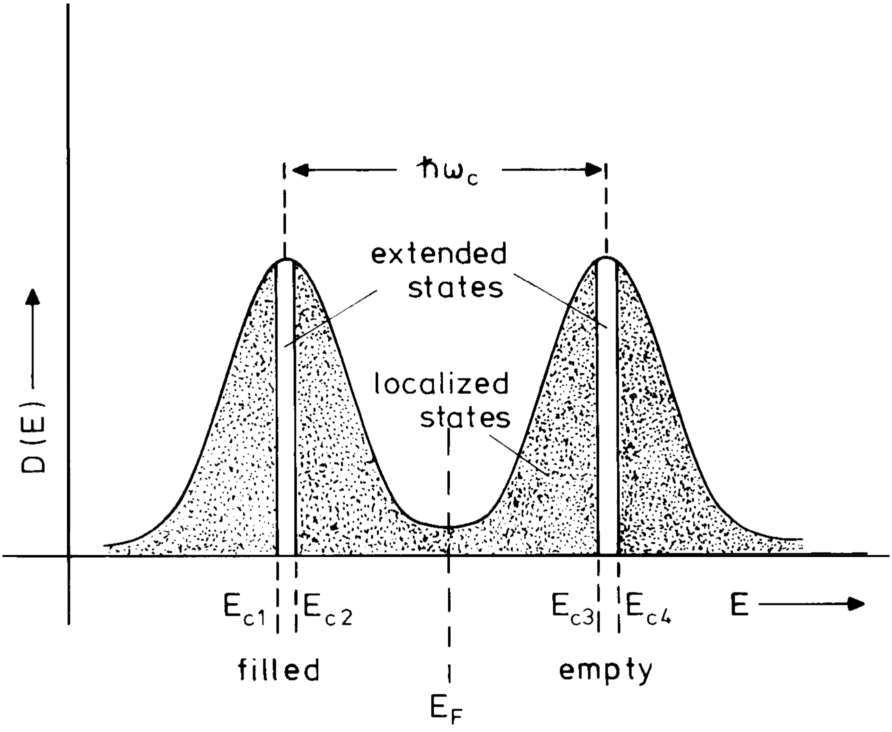


Fig. 6. Model for the broadened density of states of a 2 DEG in a strong magnetic field. Mobility edges close to the center of the Landau levels separate extended states from localized states.

4. Experimental Data

Magnetoquantum transport measurements on two-dimensional systems are known and published for more than 20 years. The first data were obtained with silicon MOSFETs and at the beginning mainly results for the conductivity σ_{xx} as a function of the carrier density (gate voltage) were analyzed. A typical curve is shown in Fig. 7. The conductivity oscillates as a function of the filling of the Landau levels and becomes zero at certain gate voltages V_g. In strong magnetic fields σ_{xx} vanishes not only at a fixed value V_g but in a range ΔV_g, and Kawaji was the first one who pointed out that some kind of immobile electrons must be introduced [22], since the conductivity σ_{xx} remains zero even if the carrier density is changed. However, no reliable theory was available for a discussion of localized electrons, whereas the peak value of σ_{xx} was well explained by calculations based on the self-consistent Born approximation and short-range scatterers which predict $\sigma_{xx} \sim (n + 1/2)$ independent of the magnetic field.

The theory for the Hall conductivity is much more complicated, and in the lowest approximation one expects that the Hall conductivity σ_{xy} deviates from the classical curve $\sigma_{xy}^0 = -\frac{n_s e}{B}$ (where n_s is the total number of electrons in the two-dimensional system per unit area) by an amount $\Delta\sigma_{xy}$ which depends mainly on the third power of the density of states at the Fermi energy [23].

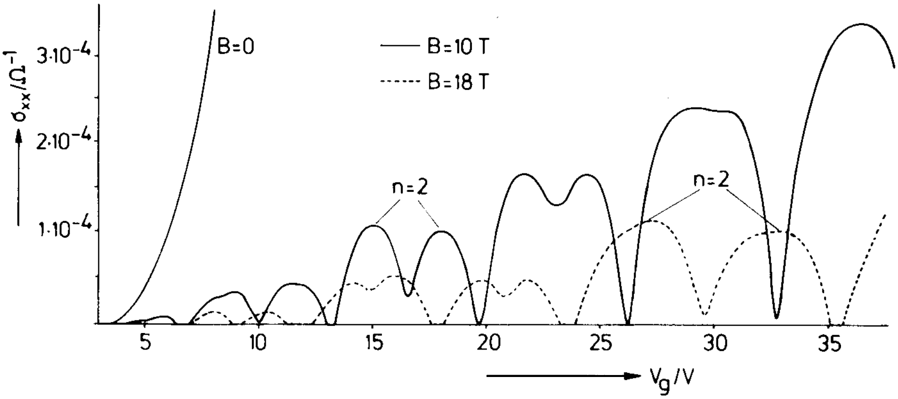


Fig. 7. Conductivity σ_{xx} of a silicon MOSFET at different magnetic fields B as a function of the gate voltage V_g.

However, no agreement between theory and experiment was obtained. Today, it is believed, that $\Delta\sigma_{xy}$ is mainly influenced by localized states, which can explain the fact that not only a positive **but** also a negative sign for $\Delta\sigma_{xy}$ is observed. Up to 1980 all experimental Hall effect data were analyzed on the basis of an incorrect model so that the quantized Hall resistance, which is already visible in the data published in 1978 [24] remained unexplained.

Whereas the conductivity σ_{xx} can be measured directly by using a Corbino disk geometry for the sample, the Hall conductivity is not directly accessible in an experiment but can be calculated from the longitudinal resistivity ϱ_{xx} and the Hall resistivity ϱ_{xy} measured on samples with Hall geometry (see Fig. 3):

$$\sigma_{xy} = \frac{\rho_{xy}}{\rho_{xx}^2 + \rho_{xy}^2}, \quad \sigma_{xx} = \frac{\rho_{xx}}{\rho_{xx}^2 + \rho_{xy}^2} \tag{16}$$

Fig. 8 shows measurements for ϱ_{xx} and ϱ_{xy} of a silicon MOSFET as a function of the gate voltage at a fixed magnetic field. The corresponding σ_{xx}- and σ_{xy} -data are calculated on the basis of Eq. (16).

The classical curve $\sigma_{xy}^0 = -\frac{n_s e}{B}$ in Fig. 8 is drawn on the basis of the incorrect model, that the experimental data should lie always below the classical curve (= fixed sign for A σ_{xy}) so that the plateau value σ_{xy} = const. (observable in the gate voltage region where σ_{xx} becomes zero) should change with the width of the plateau. Wider plateaus should give smaller values for $|\sigma_{xy}|$. The main discovery in 1980 was [23] that the value of the Hall resistance in the plateau region is not influenced by the plateau width as shown in Fig. 9. Even the aspect ratio L/W (L = length, W = width of the sample), which influences normally the accuracy in Hall effect measurements, **becomes** unimportant as shown in Fig. 10. Usually, the measured Hall resistance R_H^{exp} is always smaller than the theoretical value $R_H^{theor} = \rho_{xy}$ [26, 27]

$$R_H^{exp} = G \cdot R_H^{theor} \quad G < 1 \tag{17}$$

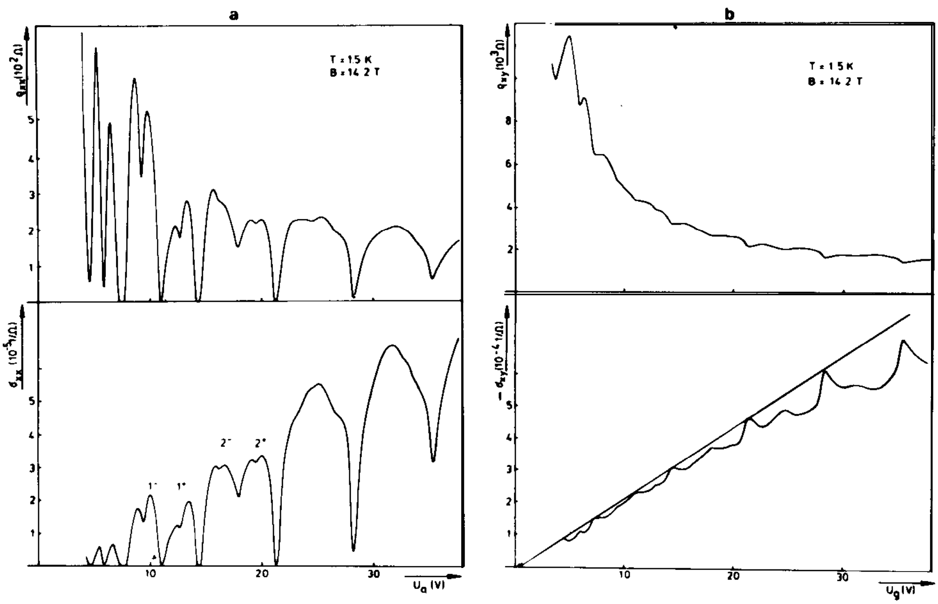


Fig. 8. Measured ρ_{xx}- and ρ_{xy}-data of a silicon MOSFET as a function of the gate voltage at B = 14.2 together with the calculated σ_{xx}- and σ_{xy}-curves.

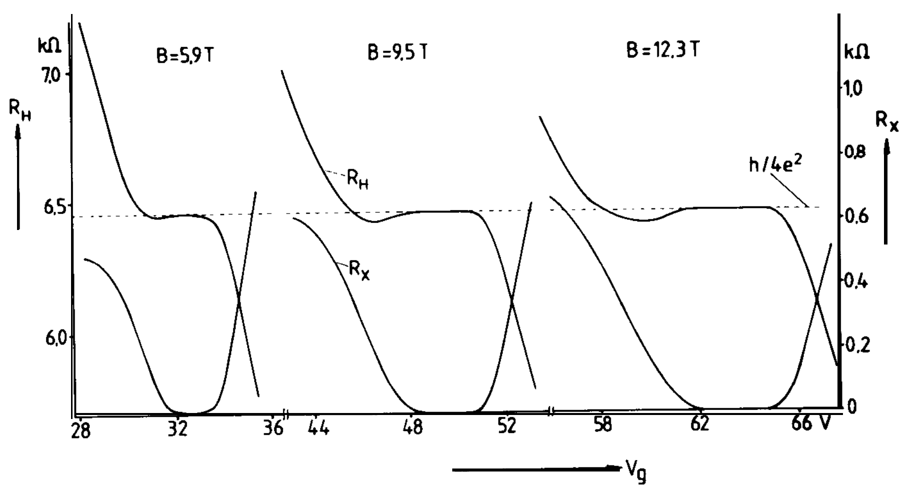


Fig. 9. Measurements of the Hall resistance R_H and the resistivity R_x as a function of the gate voltage at different magnetic field values. The plateau values $R_H = h/4e^2$ are independent of the width of the plateaus.

However, as shown in Fig. 11, the correction 1-G becomes zero (independent of the aspect ratio) if $\sigma_{xx} \rightarrow 0$ or the Hall angle θ approaches 90° ($\tan\theta = \frac{\sigma_{xy}}{\sigma_{xx}}$). This means that any shape of the sample can be used in QHE-experiments as long as the Hall angle is 90° (or $\sigma_{xx} = 0$). However, outside the plateau region ($\sigma_{xx} \sim \rho_{xx} \neq 0$) the measured Hall resistance $R_H^{exp} = \frac{U_H}{I}$ is indeed always

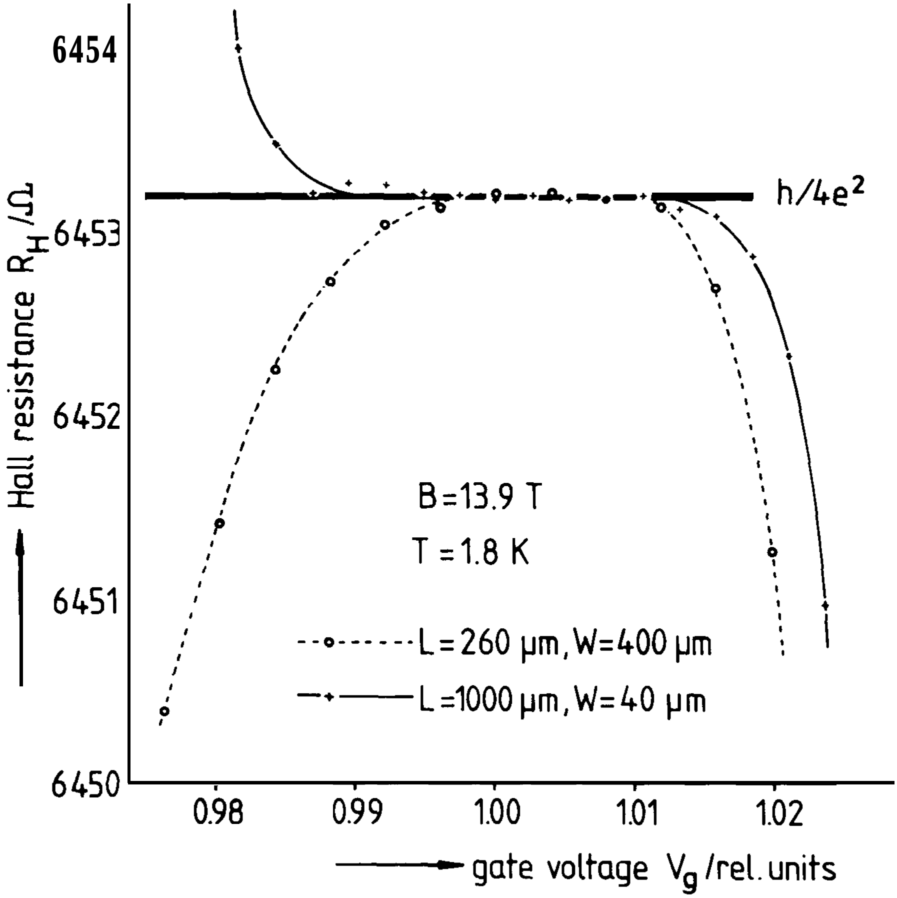


Fig. 10. Hall resistance R_H for two different samples with different aspect ratios L/W as a function of the gate voltage (B = 13.9 T).

smaller than the theoretical pxy value [28]. This leads to the experimental result that an additional minimum in R_H^{eXP} becomes visible outside the plateau region as shown in Fig. 9, which disappears if the correction due to the finite length of the sample is included (See Fig. 12). The first high-precision measurements in 1980 of the plateau value in R_H(Vg) showed already that these resistance values are quantized in integer parts of h/e^2= 25812.8 Ω within the experimental uncertainty of 3 ppm.

The Hall plateaus are much more pronounced in measurements on GaAs-Al_xGal-xAs heterostructures, since the small effective mass m* of the electrons in GaAs (m*(Si) /m* (GaAs) > 3) leads to a relatively large energy splitting between Landau levels (Eq. 4), and the high quality of the GaAs-$Al_xGa_{1-x}As$ interface (nearly no surface roughness) leads to a high mobility µ of the electrons, so that the condition µB > 1 for Landau quantizations is fulfilled already at relatively low magnetic fields. Fig. 13 shows that well-developed Hall plateaus are visible for this material already at a magnetic field strength of 4 tesla. Since a finite carrier density is usually present in heterostructures, even

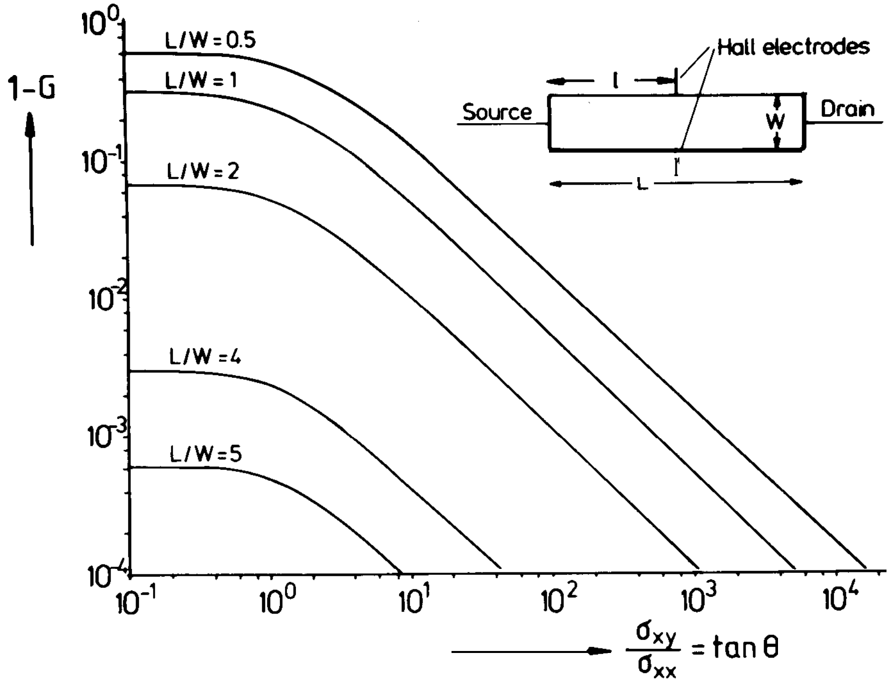


Fig. II. Calculations of the correction term G in Hall resistance measurements due to the finite length to width ratio L/W of the device (l/L = 0,5).

at a gate voltage V_g = 0V, most of the published transport data are based on measurements without applied gate voltage as a function of the magnetic field. A typical result is shown in Fig. 14. The Hall resistance $R_H = \varrho_{xy}$ increases steplike with plateaus in the magnetic field region where the longitudinal resistance ϱ_{xx} vanishes. The width of the ϱ_{xx}-peaks in the limit of zero temperature can be used for a determination of the amount of extended states and the analysis [29] shows, that only few percent of the states of a Landau level are not localized. The fraction of extended states within one Landau level decreases with increasing magnetic field (Fig. 15) but the number of extended states within each level remains approximately constant since the degeneracy of each Landau level increases proportionally to the magnetic field.

At finite temperatures ϱ_{xx} is never exactly zero and the same is true for the slope of the ϱ_{xy}-curve in the plateau region. But in reality, the slope $d\varrho_{xy}/dB$ at T<2K and magnetic fields above 8 Tesla is so small that the ϱ_{xy}-value stays constant within the experimental uncertainty of 6 10^{-8} even if the magnetic field is changed by 5 %. Simultaneously the resistivity ϱ_{xx} is usually smaller than $1 m\Omega$. However, at higher temperatures or lower magnetic fields a finite resistivity ϱ_{xx} and a finite slope $d\varrho_{xy}/dn_s$ (or $d\varrho_{xy}/dB$) can be measured. The data are well described within the model of extended states at the energy position of the undisturbed Landau level E_n and a finite density of localized states between the Landau levels (mobility gap). Like in amorphous systems, the temperature dependence of the conductivity σ_{xx} (or resistivity ϱ_{xx}) is thermally activated

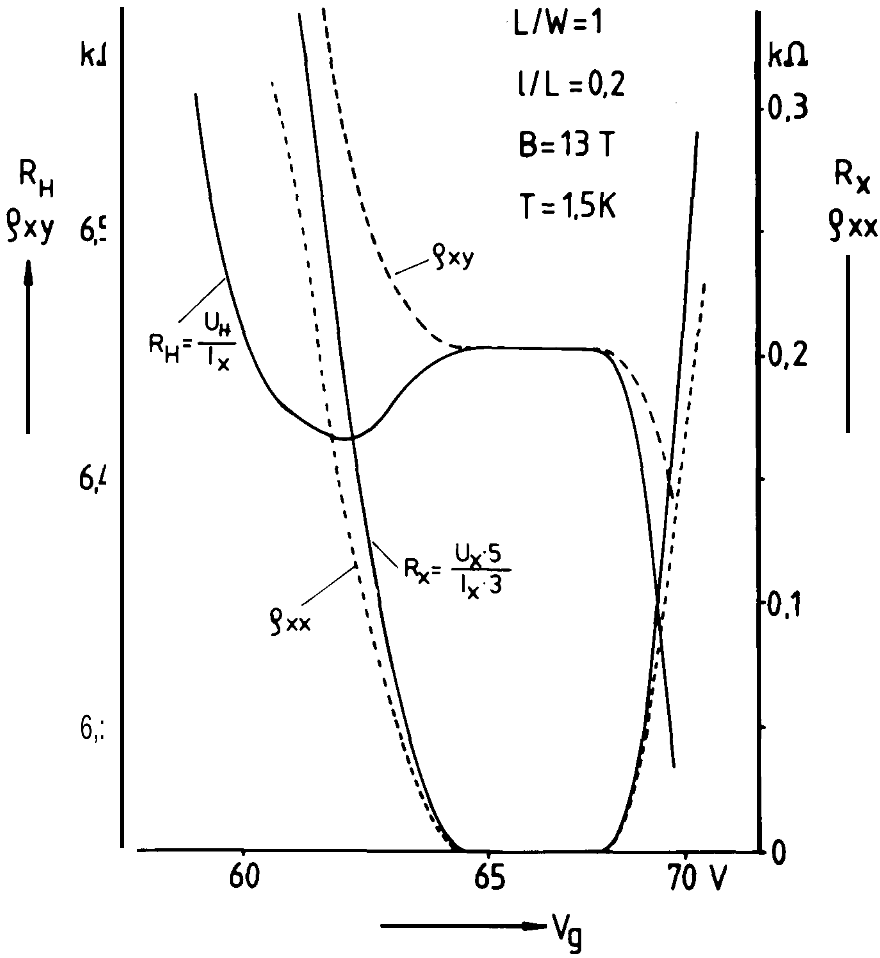


Fig. 12. Comparison between the measured quantities R_H and R_x and the corresponding resistivity components ϱ_{xy} and ϱ_{xx}, respectively.

with an activation energy E_a corresponding to the energy difference between the Fermi energy E_F and the mobility edge. The largest activation energy with a value $E_a = 1/2\hbar\omega_c$ (if the spin splitting is negligibly small and the mobility edge is located at the center E_n of a Landau level) is expected if the Fermi energy is located exactly at the midpoint between two Landau levels.

Experimentally, an activated resistivity

$$\varrho_{xx} \sim \exp\left[-(E_a/kT)\right] \tag{18}$$

is observed in a wide temperature range for different two-dimensional systems (deviations from this behaviour, which appear mainly at temperatures below 1K, will be discussed separately) and a result is shown in Fig. 16. The activation energies (deduced from these data) arc plotted in Fig. 17 for both, silicon MOSFETs and GaAs-$Al_xGa_{1-x}As$ heterostructures as a function of the

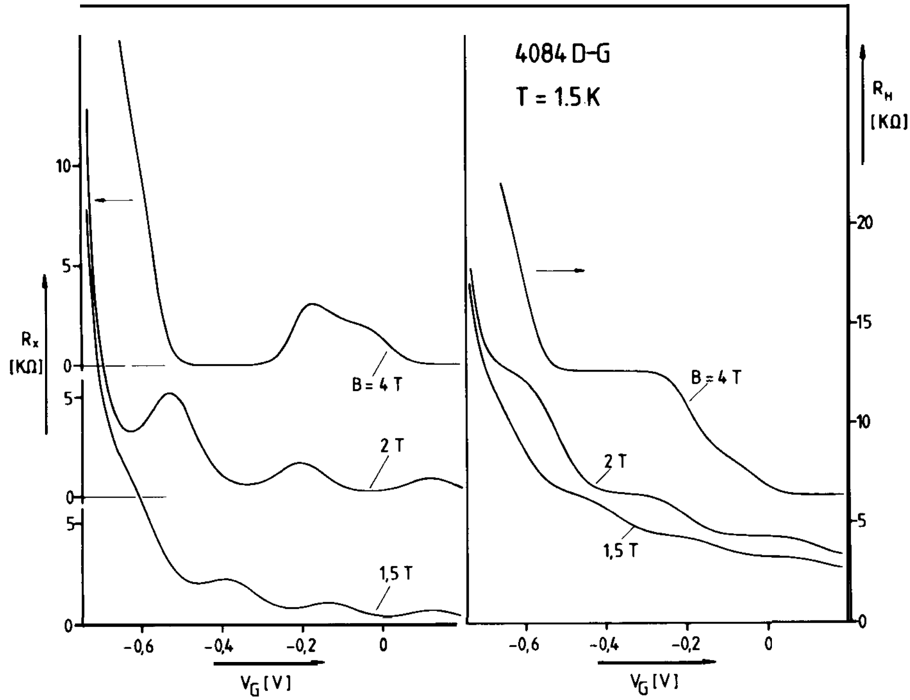


Fig. 13. Measured curves for the Hall resistance R_H and the longitudinal resistance R_x of a GaAs-$Al_xGa_{1-x}As$ heterostructure as a function of the gate voltage at different magnetic fields.

magnetic field and the data agree fairly well with the expected curve $E_a = 1/2\hbar\omega_c$. Up to now, it is not clear whether the small systematic shift of the measured activation energies to higher values originates from a temperature dependent prefactor in Eq. (18) or is a result of the enhancement of the energy gap due to many body effects.

The assumption, that the mobility edge is located close to the center of a Landau level E_n is supported by the fact that for the samples used in the experiments only few percent of the states of a Landau level are extended [29]. From a systematic analysis of the activation energy as a function of the tilling factor of a Landau level it is possible to determine the density of states D(E) [30]. The surprising result is, that the density of states (DOS) is finite and approximately constant within 60% of the mobility gap as shown in Fig. (18). This background DOS depends on the electron mobility as summarized in Fig. (19).

An accurate determination of the DOS close to the center of the Landau level is not possible by this method since the Fermi energy becomes temperature dependent if the DOS changes drastically within the energy range of 3kT. However, from an analysis of the capacitance C as a function of the Fermi energy the peak value of the DOS and its shape close to E_n can be deduced [31, 32].

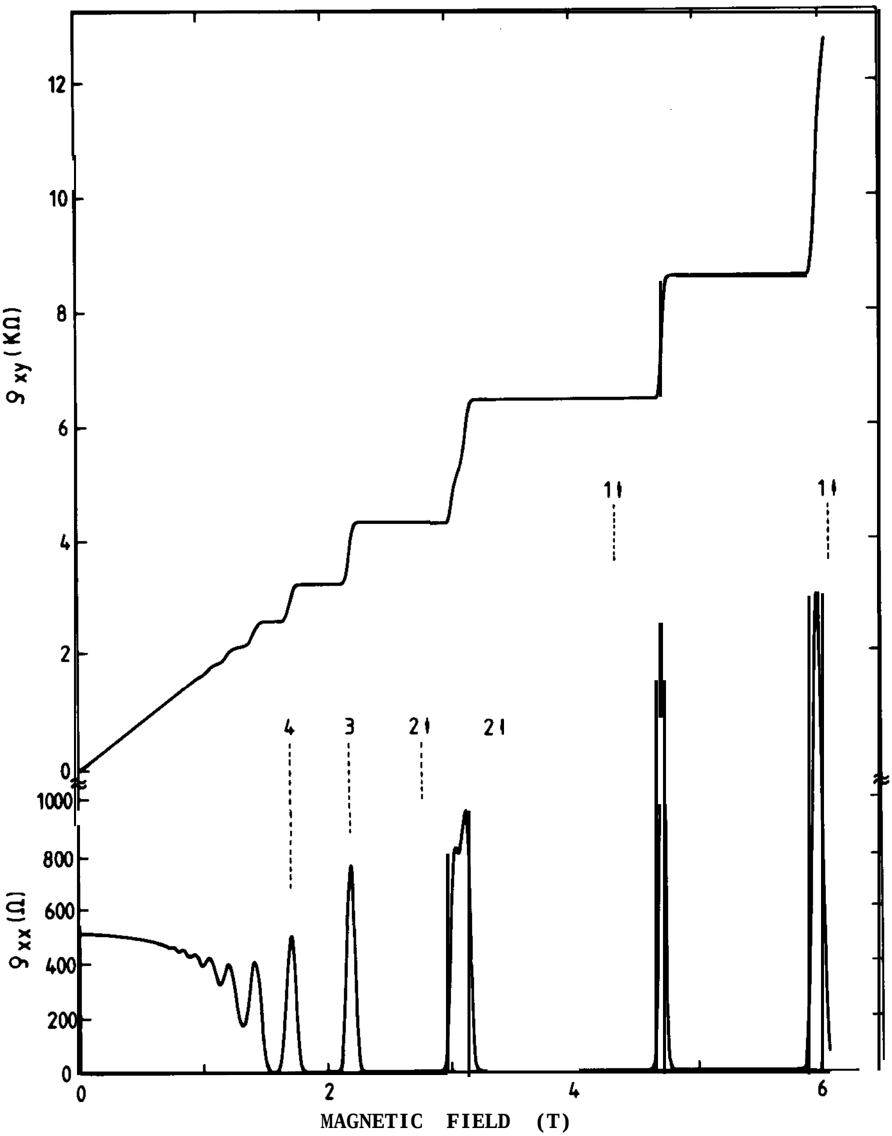


Fig. 14. Experimental curves for the Hall resistance $R_H = \varrho_{xy}$ and the resistivity $\varrho_{xx} \sim R_x$ of a heterostructure as a function of the magnetic field at a fixed carrier density corresponding to a gate voltage V_g, = 0V. The temperature is about 8mK.

This analysis is based on the equation

$$\frac{1}{C} = \frac{1}{e^2 \cdot D\,(E_F)} + \text{const.} \tag{19}$$

The combination of the different methods for the determination of the DOS leads to a result as shown in Fig. (20). Similar results are obtained from other experiments, too [33, 34] but no theoretical explanation is available.

If one assumes that only the occupation of extended states influences the Hall effect, than the slope $d\varrho_{xy}/dn_s$ in the plateau region should be dominated

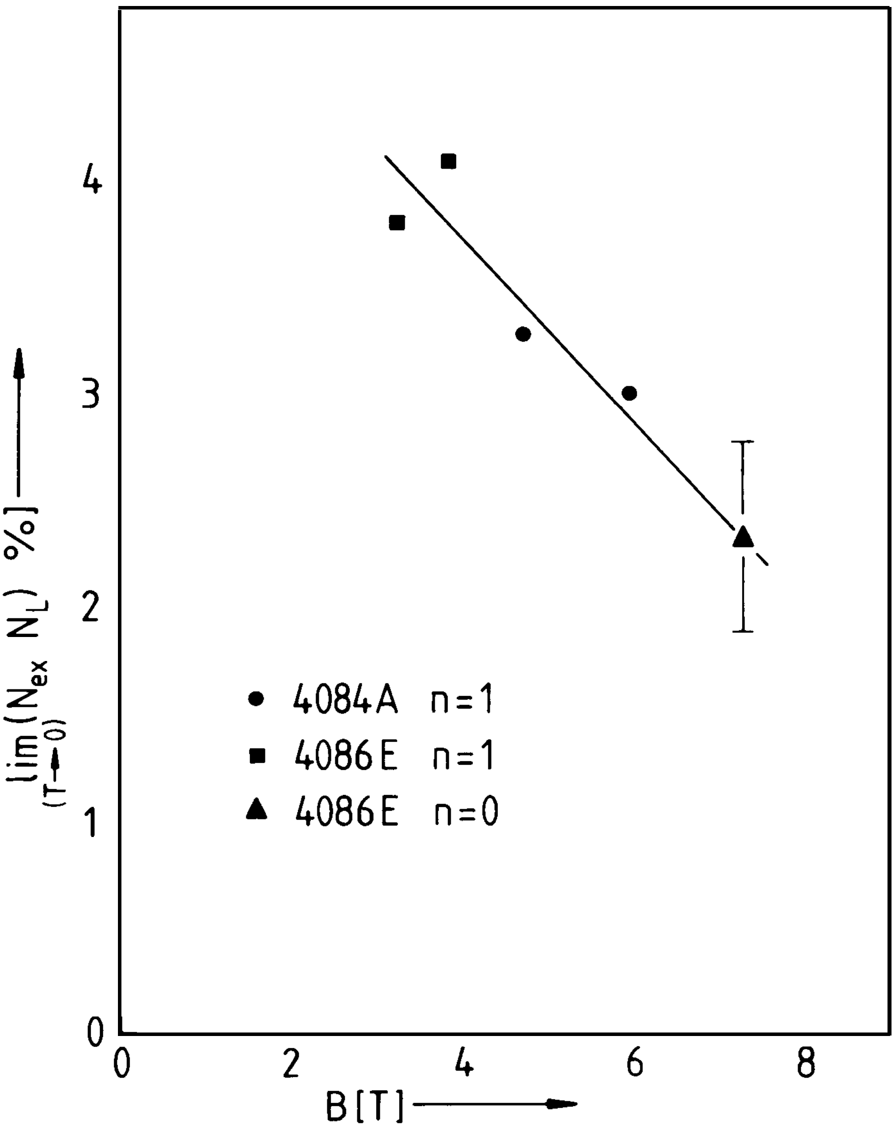


Fig. 15. Fraction of extended states relative to the number of states of one Landau level as a function of the magnetic field.

by the same activation energy as found for $\varrho_{xx}(T)$. Experimentally [35], a one to one relation between the minimal resistivity ϱ_{xx}^{min} at integer filling factors and the slope of the Hall plateau has been found (Fig. 21) so that the flatness of the plateau increases with decreasing resistivity, which means lower temperature or higher magnetic fields.

The temperature dependence of the resistivity for Fermi energies within the mobility gap deviates from an activated behaviour at low temperatures, typically at T<lK. Such deviations are found in measurements on disordered

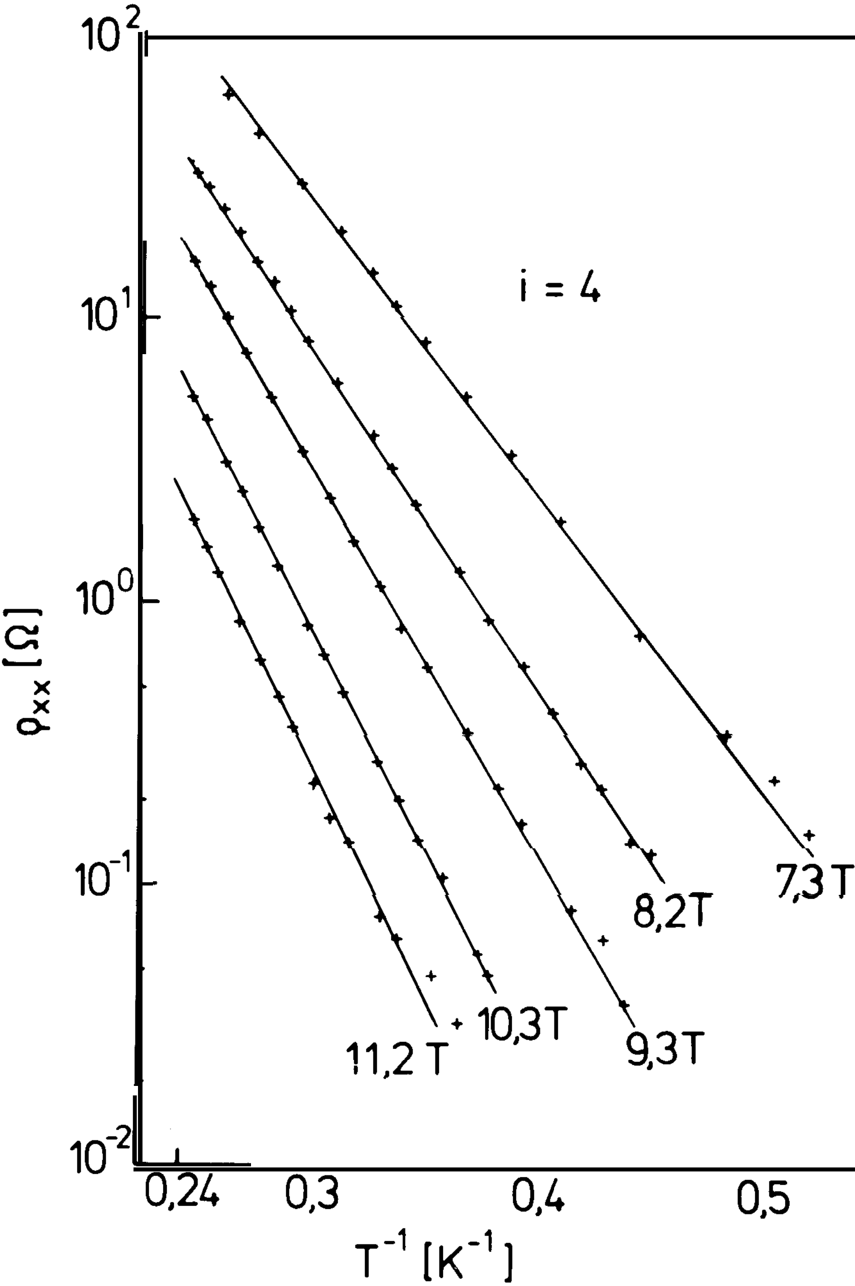


Fig 16. Thermally activated resistivity ϱ_{xx} at a filling factor i = 4 for a silicon MOSFET at different magnetic field values.

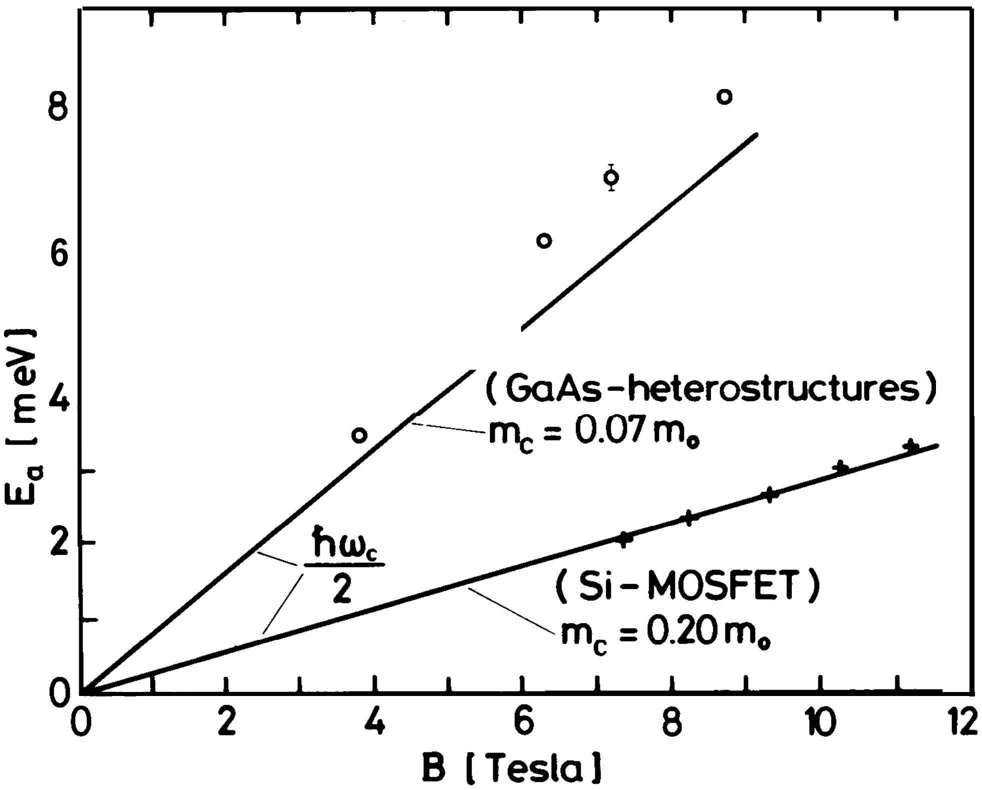


Fig. 17. Measured activation energies at filling factors i = 2 (GaAs heterostructure) or i = 4 (Si-MOSFET) as a function of the magnetic field. The data are compared with the energy $0{,}5\hbar\omega_c$.

systems, too, and are interpreted as variable range hopping. For a two-dimensional system with exponentially localized states a behaviour

$$\rho_{xx} \sim \exp\left[-(T_0/T)^{1/3}\right] \tag{20}$$

is expected. For a Gaussian localization the following dependence is predicted [36, 37]

$$\rho_{xx} \sim \frac{1}{T}\exp\left\{-(T_0/T)^{1/2}\right\} \tag{21}$$

The analysis of the experimental data demonstrates (Fig. 22) that the measurements are best described on the basis of Eq. (2 1). The same behaviour has been found in measurements on another two-dimensional system, on InP-InGaAs heterostructures [38].

The contribution of the variable range hopping (VRH) process to the Hall effect is negligibly small [39] so that experimentally the temperature dependence of $d\rho_{xy}/dn_s$ remains thermally activated even if the resistivity ρ_{xx} is dominated by VRH.

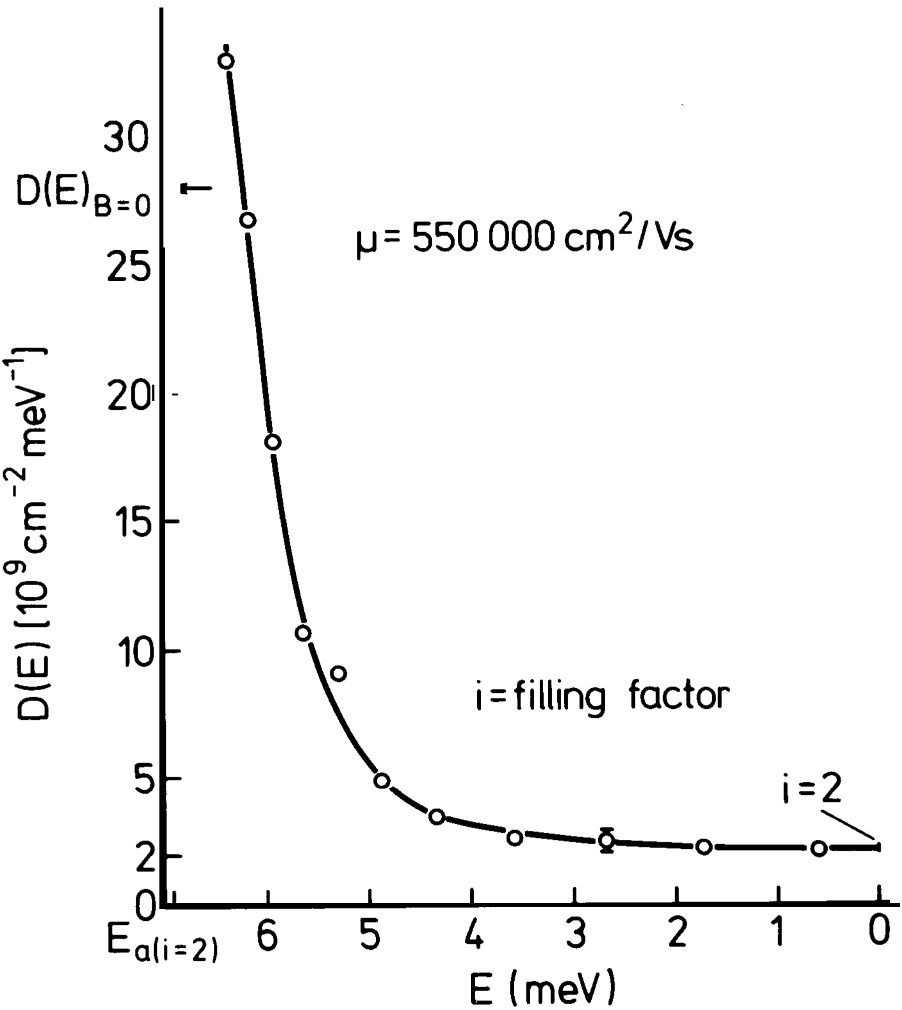


Fig. 18. Measured density of states (deduced from an analysis of the activated resisitivity) as a function of the energy relative to the center between two Landau levels (GaAs-heterostructure).

The QHE breaks down if the Hall field becomes larger than about $E_H = 60V/cm$ at magnetic fields of 5 Tesla.

This corresponds to a classical drift velocity $v_D = \frac{E_H}{B} \approx 1200 m/s$. At the critical Hall field E_H (or current density j) the resistivity increases abruptly by orders of magnitude and the Hall plateau disappears. This phenomenon has been observed by different authors for different materials [40-47]. A typical result is shown in Fig. 23. At a current density of $j_c = 0{,}5\ A/m$ the resistivity ϱ_{xx} at the center of the plateau (filling factor i = 2) increases drastically. This instability, which develops within a time scale of less than 100 ns seems to originate from a runaway in the electron temperature but also other mechanism like electric field dependent delocalization, Zener tunneling or emission of

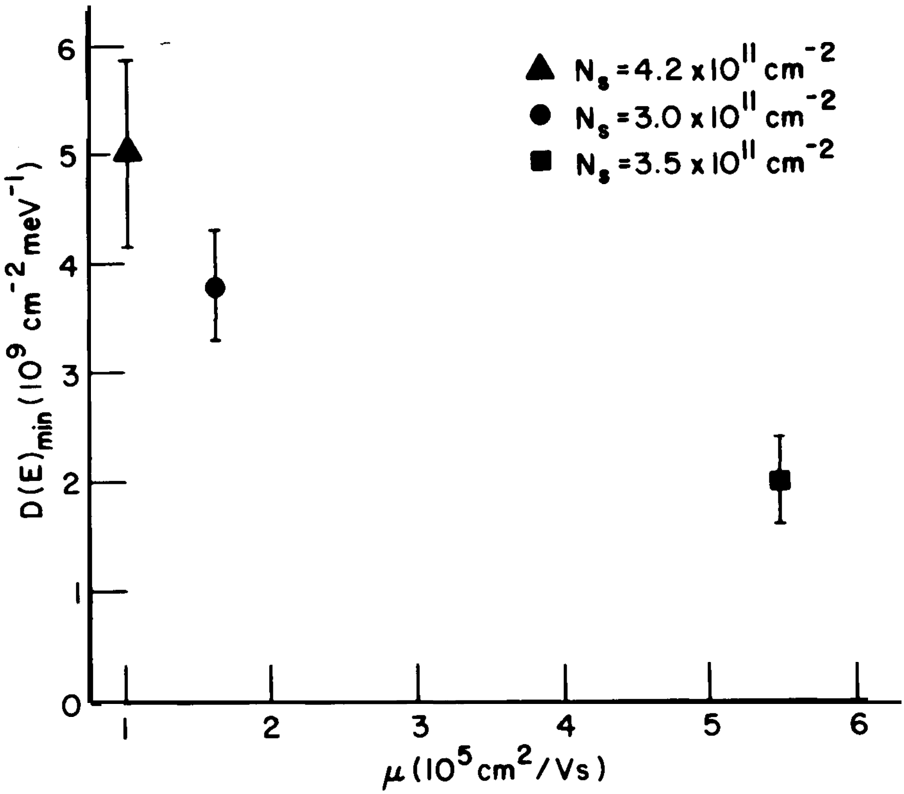


Fig. 19. Background density of states as a function of the mobility of the device.

acoustic phonons, if the drift velocity exceeds the sound velocity, can be used for an explanation [48-50].

Fig. 23 shows that ϱ_{xx} increases already at current densities well below the critical value j_c which may be explained by a broadening of the extended state region and therefore a reduction in the mobility gap AE. If the resistivity ϱ_{xx} is thermally activated and the mobility gap changes linearly with the Hall field (which is proportional to the current density j) then a variation

$$\ln \varrho_{xx} \sim j$$

is expected. Such a dependence is seen in Fig. 24 but a quantitative analysis is difficult since the current distribution within the sample is usually inhomogenious and the Hall field, calculated from the Hall voltage and the width of the sample, represents only a mean value. Even for an ideal two-dimensional system an inhomogenious Hall potential distribution across the width of the sample is expected [51-53] with an enhancement of the current density close to the boundaries of the sample.

The experimental situation is still more complicated as shown in Fig. 25. The potential distribution depends strongly on the magnetic field. Within the plateau region the current path moves with increasing magnetic field across the

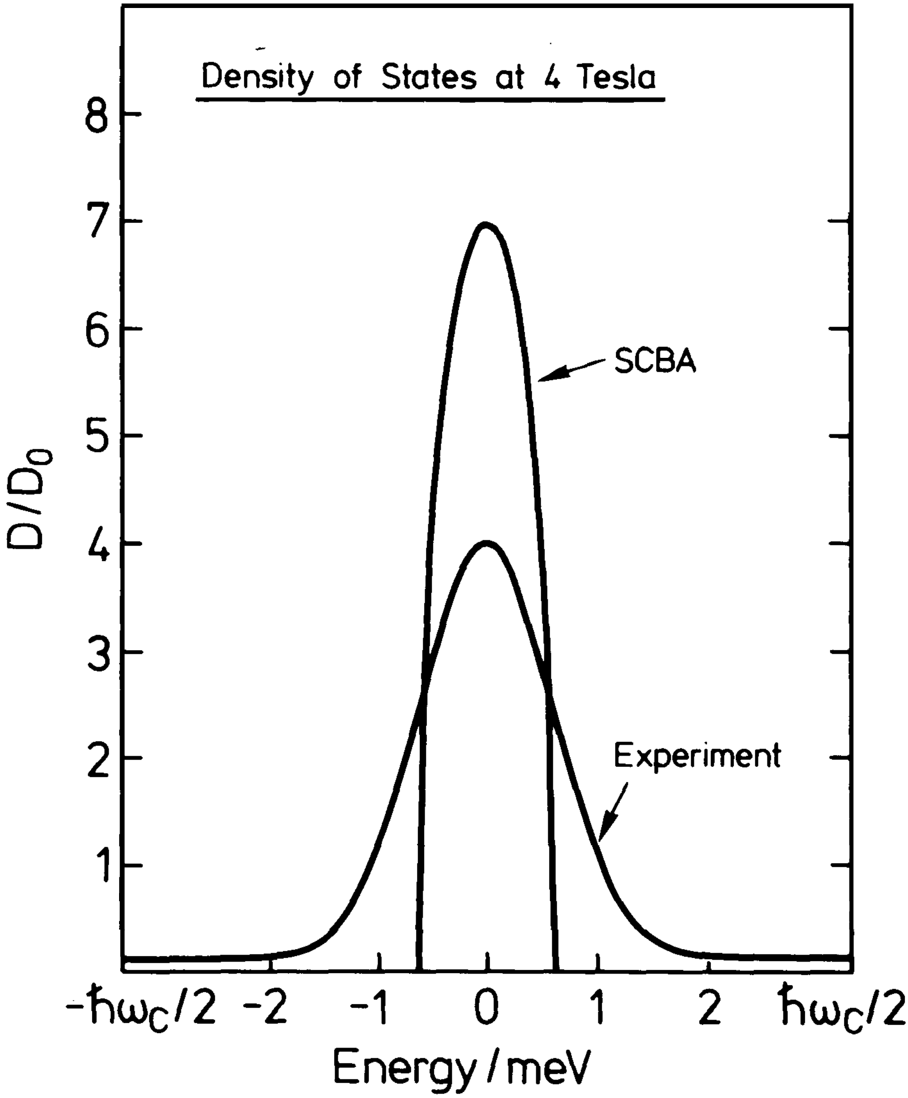


Fig. 20. Experimentally deduced density of states of a GaAs heterostructure at B = 4T compared with the calculated result based on the self-consistent Born approximation (SCBA).

width of the sample from one edge to the other one. A gradient in the carrier density within the two-dimensional system seems to be the most plausible explanation but in addition an inhomogeneity produced by the current itself may play a role. Up to now, not enough microscopic details about the two-dimensional system are known so that at present a microscopic theory, which describes the QHE under real experimental conditions, is not available. However, all experiments and theories indicate that in the limit of vanishing resistivity ϱ_{xx} the value of the quantized Hall resistance depends exclusively on fundamental constants. This leads to a direct application of the QHE in metrology.

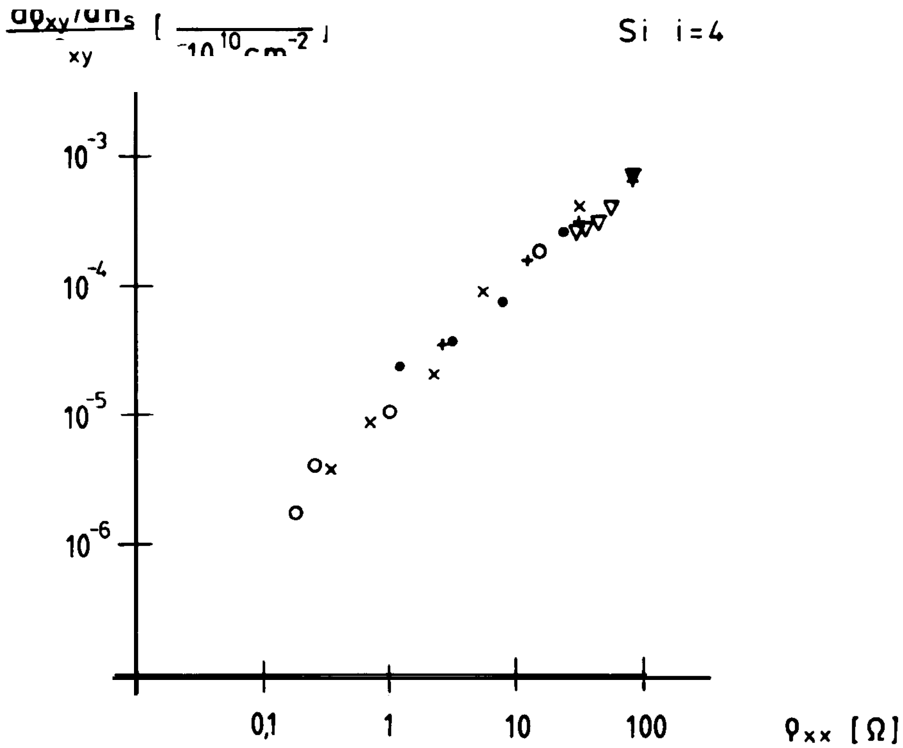


Fig. 21. Relation between the slope of the Hall plateaus $d\varrho_{xy}/dn_s$ and the corresponding ϱ_{xx}-value at integer filling factors.

5. Application of the Quantum Hall Effect in Metrology

The applications of the Quantum Hall Effect are very similar to the applications of the Josephson-Effect which can be used for the determination of the fundamental constant h/e or for the realization of a voltage standard. In analogy, the QHE can be used for a determination of h/e^2 or as a resistance standard. [54].

Since the inverse fine structure constant α^{-1} is more or less identical with h/e^2(the proportional constant is a fixed number which includes the velocity of light), high precision measurements of the quantized Hall resistance are important for all areas in physics which are connected with the finestructure constant.

Experimentally, the precision measurement of *a* is reduced to the problem of measuring an electrical resistance with high accuracy and the different methods and results are summarized in the Proceedings of the 1984 Conference on Precision Electromagnetic Measurements (CPEM 84) [55]. The mean value of measurements at laboratories in three different countries is

$$\alpha^{-1} = 137{,}035988 \pm 0.00002$$

The internationally recommended value (1973) is

$$\alpha^{-1} = 137{,}03604 \pm 0.00011$$

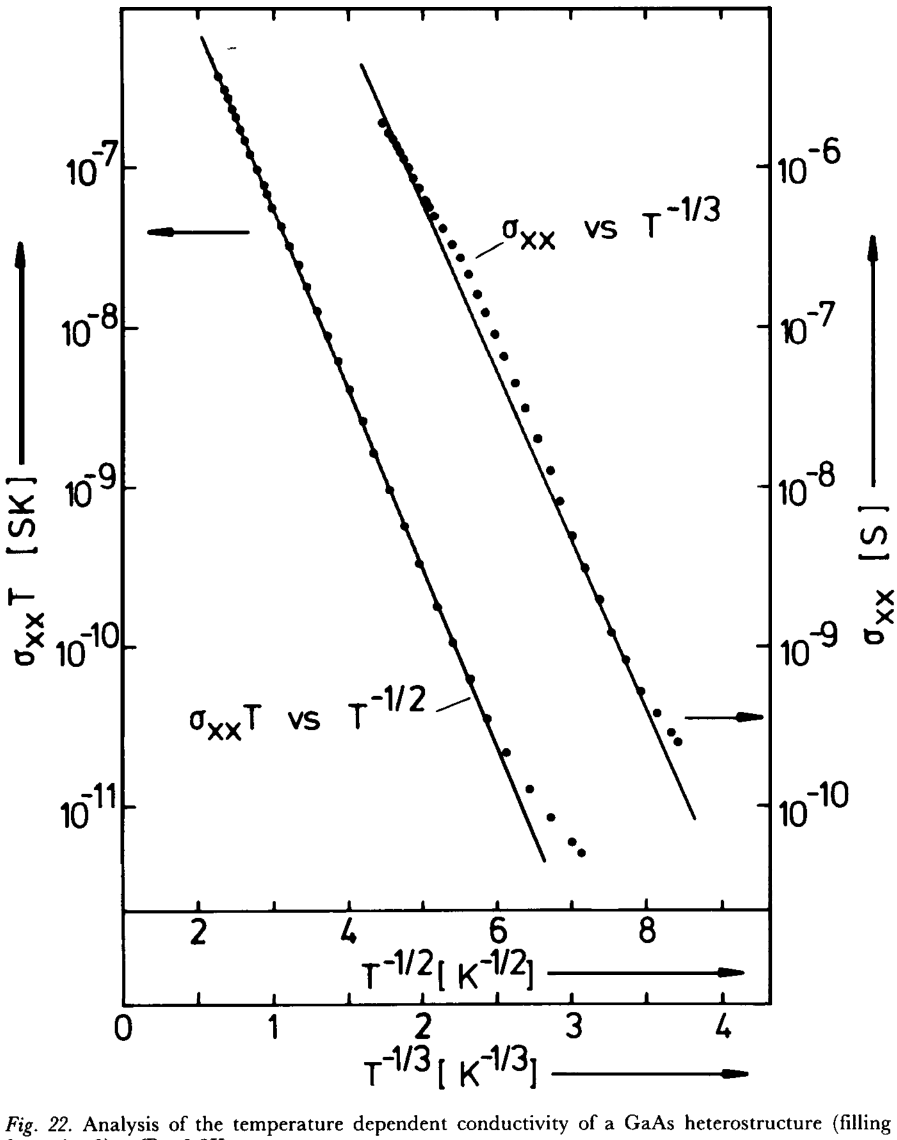


Fig. 22. Analysis of the temperature dependent conductivity of a GaAs heterostructure (filling factor i = 3) at T < 0,2K.

and the preliminary value for the tinestructure constant based on a new least square adjustment of fundamental constants (1985) is

$$\alpha^{-1} = 137{,}035991 \pm 0.000008$$

Different groups have demonstrated that the experimental result is within the experimental uncertainty of less than $3.7 \cdot 10^{-8}$ independent of the material (Si, GaAs, $In_{0.53}Ga_{0.47}As$) and of the growing technique of the devices (MBE or MOCVD) [56]. The main problem in high precision measurements of α is-at present-the calibration and stability of the reference resistor. Fig. 26 shows the drift of the maintained 1Ω-resistor at different national laboratories. The

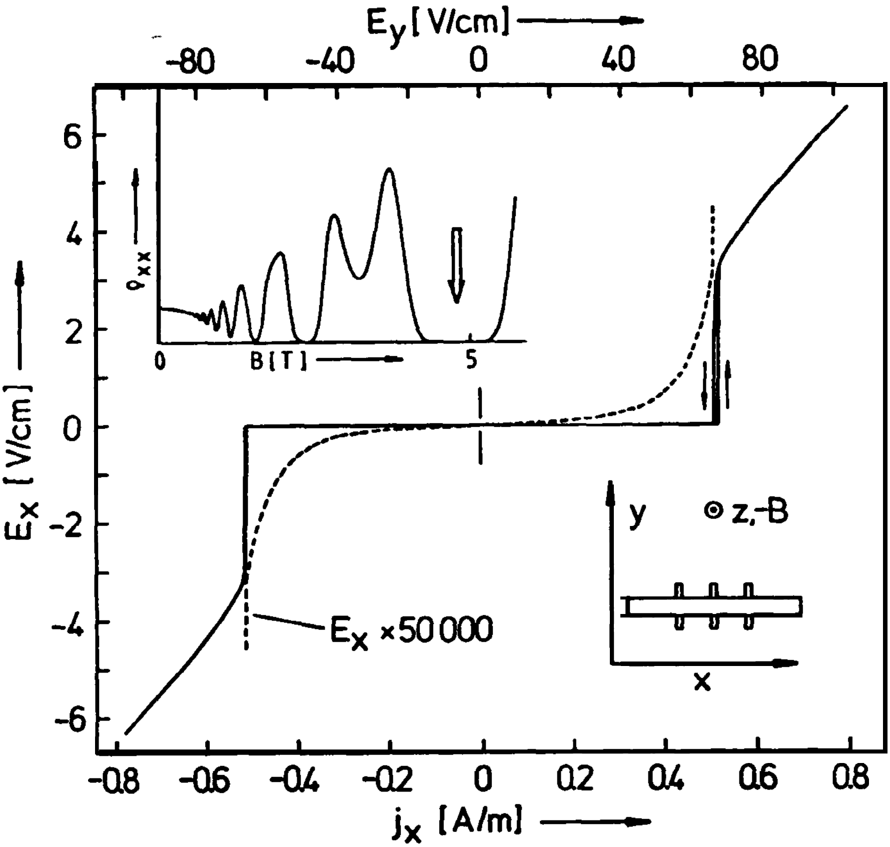


Fig. 23. Current-voltage characteristic of a GaAs-$Al_xGa_{1-x}As$ heterostructure at a filling factor i = 2 (T = 1,4K). The device geometry and the ϱ_{xx}(B)-curve are shown in the inserts.

very first application of the QHE is the determination of the drift coefficient of the standard resistors since the quantized Hall resistance is more stable and more reproducible than any wire resistor. A nice demonstration of such an application is shown in Fig. 27. In this experiment the quantized Hall resistance R_H has been measured at the "Physikalisch Technische Bundesanstalt" relative to a reference resistor R_R as a function of time. The ratio R_H / R_R changes approximately linearly with time but the result is independent of the QHE-sample. This demonstrates that the reference resistor changes its value with time. The one standard deviation of the experimental data from the mean value is only $2.4 \cdot 10^{-8}$ so that the QHE can be used already today as a relative standard to maintain a laboratory unit of resistance based on wire-wound resistors. There exists an agreement that the QHE should be used as an absolute resistance standard if three independent laboratories measure the same value for the quantized Hall resistance (in SI-units) with an uncertainty of less than $2 \cdot 10^{-7}$. It is expected that these measurements will be finished until the end of 1986.

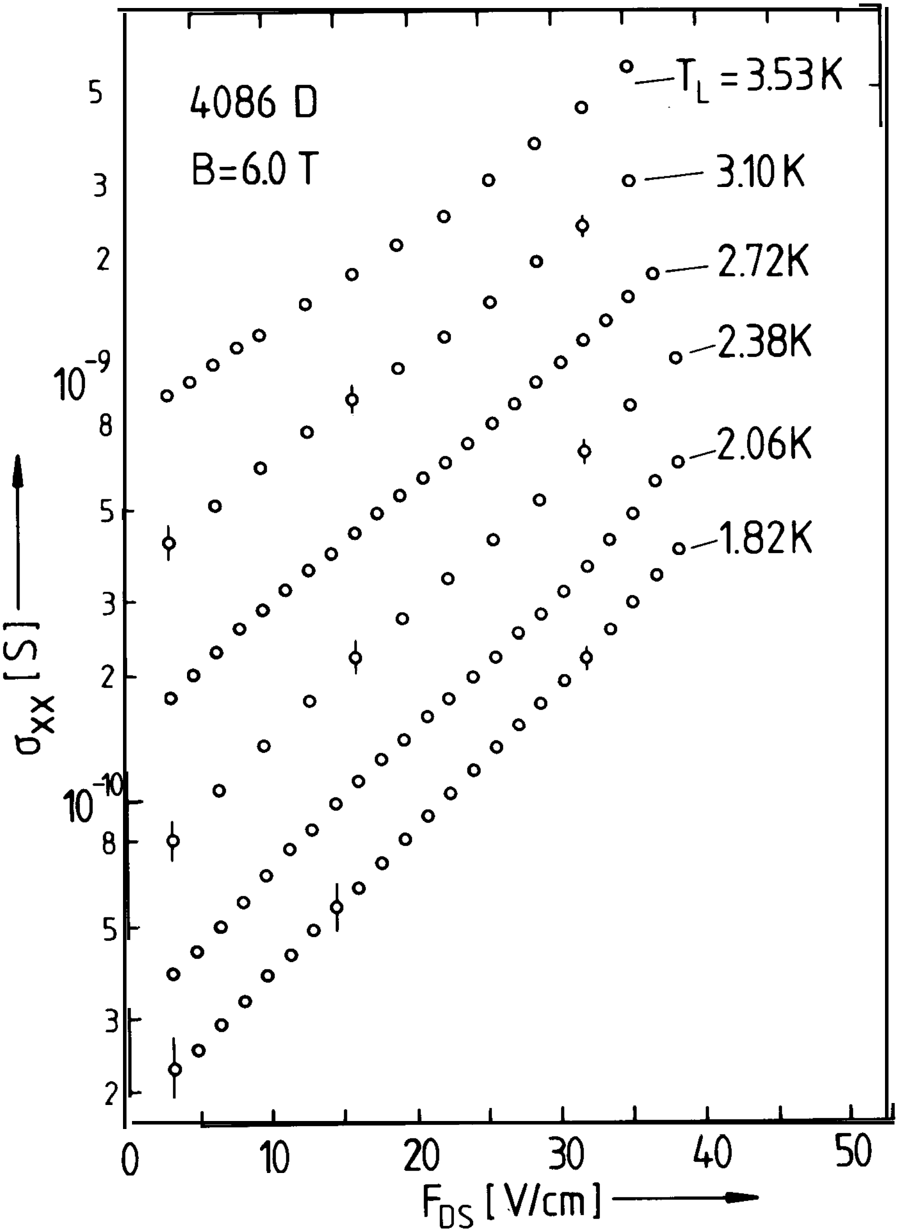


Fig. 24. Nonohmic conductivity σ_{xx} of a GaAs heterostructure at different temperatures T_L (filling factor i = 2). An instability is observed at source-drain fields larger than 40 V/cm.

Acknowledgements

The publicity of the Nobel Prize has made clear that the research work connected with the Quantum Hall Effect was so successful because a tremendous large number of institutions and individuals supported this activity. I would like to thank all of them and I will mention by name only those scientists who supported my research work at the time of the discovery of the QHE in

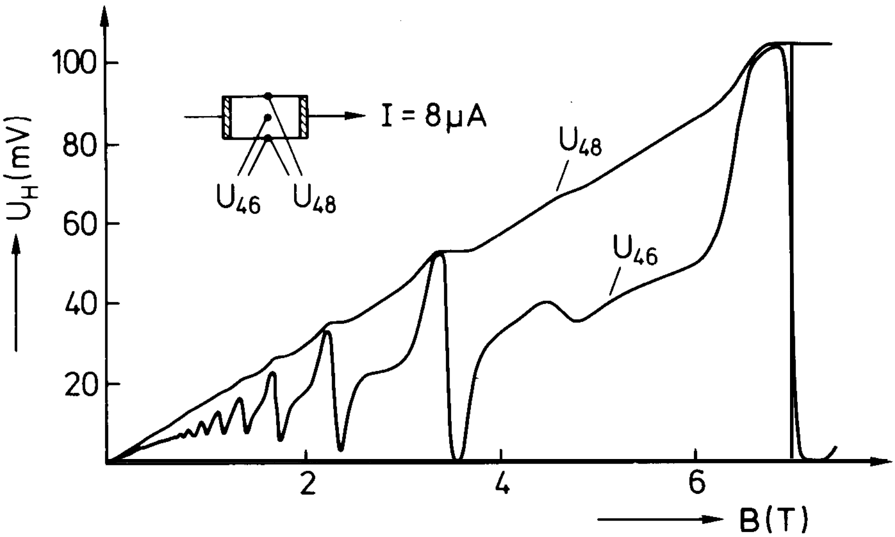


Fig. 25. Measured Hall potential distribution of a GaAs heterostructure as a function of the magnetic field.

1980. Primarily, I would like to thank G. Dorda (Siemens Forschungslaboratorien) and M. Pepper (Cavendish Laboratory, Cambridge) for providing me with high quality MOS-devices. The continuous support of my research work by G. Landwehr and the fruitful discussions with my coworker, Th. Englert, were essential for the discovery of the Quantum Hall Effect and are greatfully acknowledged.

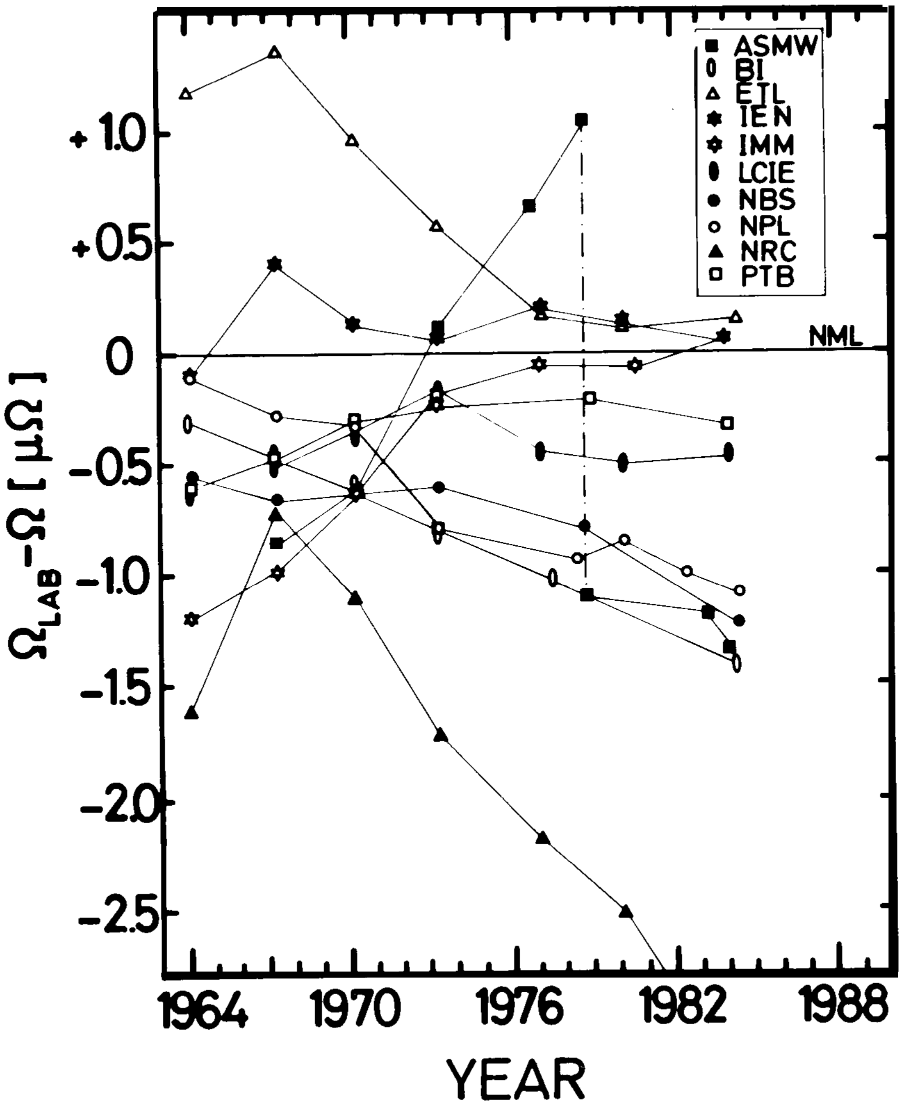


Fig. 26. Time dependence of the I Ω standard resistors maintained at the differcnt national laboratories.

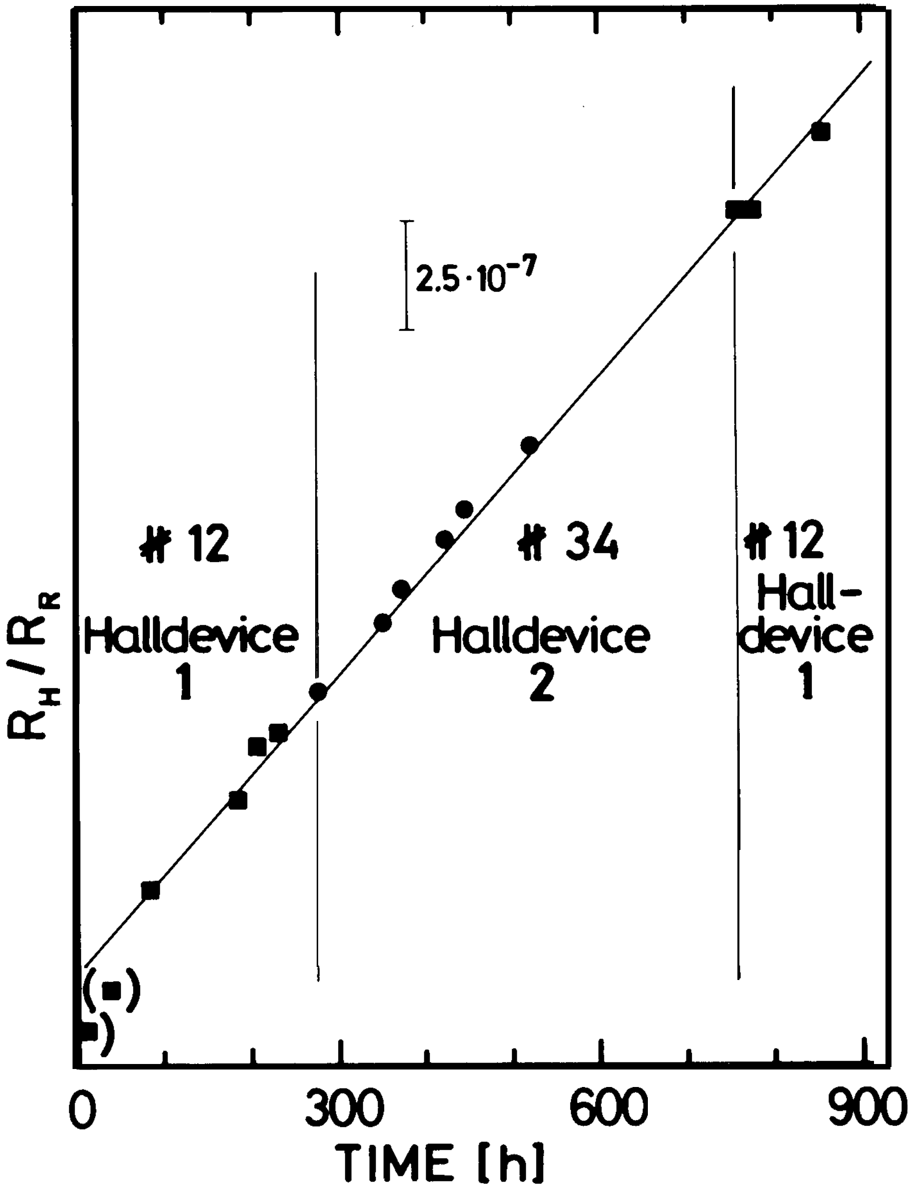


Fig. 27. Ratio R_H/R_R between the quantized Hall resistance R_H and a wire resistor R_R as a function of time, The result is time dependent but indcpcndent of the Hall device used in the experiment.

REFERENCES

[l] A.B. Fowler, F.F. Fang, W.E. Howard and P.J. Stiles Phys. Rev. Letter 16, 901 (1966)

[21 For a review see: Proceedings of the Int. Conf. on Electronic Properties of Two-Dimensional Systems, Surf. Sci. 58, (1976), 73 (1978), 98 (1980, 113 (1982), ***142* *(1984)***

[31 F. Stern and W.E. Howard, Phys. Rev. 163, 816 (1967)

[41 T. Ando, J. Phys. Soc., Jpn. 51, 3893 (1982)

[51 J.F. Koch, Festkörperprobleme (Advances in Solid State Physics), H.J. Queisser, Ed. (Pergamon-Vieweg, Braunschweig, 1975) Vol. XV, p. 79

[61 T. Mimura, Surf. Science ***113, 454*** (1982)

[71 R.B. Laughlin, Surface Science ***113, 22*** (1982)

[81 R. Kubo, S.J. Miyake and N. Hashitsume, in Solid State Physics, Vol. 17, 269 (1965). F. Seitz and D. Turnball, Eds., (Academic Press, New York, 1965)

[91 T. Ando, J. Phys. Soc. Jpn. 37, 1233 (1974)

[10] R. B. Laughlin in Springer Series in Solid State Sciences 53, p. 272, G. Bauer, F. Kuchar and H. Heinrich, Eds. (Springer Verlag, 1984)

[11] R.E. Prange, Phys. Rev. B 23, 4802 (1981)

[12] H. Aoki and T. Ando, Solid State Commun. 38, 1079 (1981)

[13] J. T. Chalker, J. Phys. C 16, 4297 (1983)

[14] W. Brenig, Z. Phys. .50B, 305 (1983)

[15] A. Mac Kinnon, L. Schweitzer and B. Kramer, Surf. Sci. ***142,*** 189 (1984)

[16] T. Ando, J. Phys. Soc. Jpn. 52, 1740 (1983)

[17] L. Schweitzer, B. Kramer and A. Mac Kinnon, J. Phys. C 17, 4111 (1984)

[18] H. Aoki and T. Ando, Phys. Rev. Letters 54, 831 (1985)

[19] E. Abrahams, P.W. Anderson, D.C. Licciardello and T.V. Ramakrishnan, Phys. Rev. Letters 42, 673 (1979)

[20] G.A. Baraff and D.C. Tsui, Phys. Rev. B 24, 2274 (1981)

[21] T. Toyoda, V. Gudmundsson and Y. Takahashi, Phys. Letters ***102A,*** 130 (1984)

[22] S. Kawaji and J, Wakabayashi, Surf. Schi. 58, 238 (1976)

[23] S. Kawaji, T. Igarashi and J. Wakabayashi, Progr. in Theoretical Physics 57, 176 (1975)

[24] T. Englert and K. v. Klitzing, Surf. Sci. 73, 70 (1978)

[25] K. v. Klitzing, G. Dorda and M. Pepper, Phys. Rev. Letters 45, 494 (1980)

[26] K. v. Klitzing, H. Obloh, G. Ebert, J, Knecht and K. Ploog, Prec. Measurement and Fundamental Constants II, B.N. Taylor and W.D. Phillips, Eds., Natl. Burl. Stand. (U.S.), Spec. Publ. ***617,*** (1984) p. 526

[27] R.W. Rendell and S.M. Girvin, Prec. Measurement and Fundamental Constants II, B.N. Taylor and W.D. Phillips. Eds. Natl. Bur. Stand. (U.S.), Spec. Publ. ***617,*** (1984) p. 557

[28] K. v. Klitzing, Festkörperprobleme (Advances in Solid State Physics), XXI, 1 (1981), J. Treusch, Ed., (Vieweg, Braunschweig)

[29] G. Ebert, K. v. Klitzing, C. Probst and K. Ploog, Solid State Commun. 44, 95 (1982)

[30] E. Stahl, D. Weiss, G. Weimann, K. v. Klitzing and K. Ploog, J. Phys. C ***18,*** L 783 (1985)

[31] T.P. Smith, B.B. Goldberg, P.J. Stiles and M. Heiblum, Phys. Rev. B 32, 2696 (1985)

[32] V. Mosser, D. Weiss, K. v. Klitzing, K. Ploog and G. Weimann to be published in Solid State Commun.

[33] E. Gornik, R. Lassnig, G. Strasser, H.L. Störmer, A.C. Gossard and W. Wiegmann, Phys. Rev. Lett. 54, 1820 (1985)

[34] J.P. Eisenstein, H.L. Störmer, V. Navayanamurti, A.Y. Cho and A.C. Gossard, Yamada Conf. XIII on Electronic Properties of Two-Dimensional Systems, p. 292 (1985)

[35] B. Tausendfreund and K. v. Klitzing, Surf. Science 142, 220 (1984)
[36] M. Pepper, Philos, Mag. 37, 83 (1978)
[37] Y. Ono, J. Phys. Soc. Jpn. 51, 237 (1982)
[38] Y. Guldner, J.P. Hirtz, A. Briggs, ,J.P. Vieren, M. Voos and M. Razeghi, Surf. Science **142, 179 (1984)**
[39] K.I. Wysokinski and W. Brenig, Z. Phys. B - Condensed Matter 54, 11 (1983)
[40] G. Ebert and K. v. Klitzing, J. Phys. C 5441 (1983)
[41] M.E. Cage, R.F. Dziuba, B.F. Field, E.R. Williams, S.M. Girvin, AC. Gossard, D.C. Tsui and R.J. Wagner, Phys. Rev. Letters 51, 1374 (1983)
[42] F. Kuchar, G. Bauer, G. Weimann and H. Burkhard, Surf. Science 142, 196 (1984)
[43] H.L. Störmer, A.M. Chang, D.C. Tsui and J.C.M. Hwang, Proc. 17th ICPS, San Francisco 1984
[44] H. Sakaki, K. Hirakawa, J. Yoshino, S.P. Svensson, Y. Sekiguchi, T. Hotta and S. Nishii, Surf. Science **142, 306 (1984)**
[45] K v. Klitzing, G. Ebert, N. Kleinmichel, H. Obloh, G. Dorda and G. Weimann, Proc. 17th ICPS, San Francisco 1984
[46] F. Kuchar, R. Meisels, G. Weimann and H. Burkhard, Proc. 17th ICPS, San Francisco 1984
[47] V.M. Pudalov and S.G. Semenchinsky, Solid State Commun. 51, 19 (1984)
[48] P. Streda and K. v. Klitzing, J. Phys. C 17, L 483 (1984)
[49] O. Heinonen, P.L. Taylor and S.M. Girvin, Phys. Rev. B 30, 3016 (1984)
[50] S.A. Trugman, Phys. Rev. B 27, 7539 (1983)
[51] A.H. Mac Donold, T.M. Rice and W.F. Brinkman, Phys. Rev. B 28, 3648 (1983)
[52] O. Heinonen and P.L. Taylor, Phys. Rev. B 32, 633 (1985)
[53] J. Riess, J. Phys. C **17,** L 849 (1984)
[54] K. v. Klitzing and G. Ebert, Metrologia 21, 11 (1985)
[55] High-precision measurements of the quantized Hall resistance are summarized in: IEEE Trans. Instrum. Meas. IM 34, pp. 301-327
[56] F. Delahaye, D. Dominguez, F. Alexandre, J. P. Andre, J. P. Hirtz and M. Razeghi, Metrologia 22, 103 (1986)

Physics 1986

ERNST RUSKA

for his fundamental work in electron optics, and for the design of the first electron microscope

GERD BINNIG and HEINRICH ROHRER

for their design of the scanning tunneling- microscope

THE NOBEL PRIZE FOR PHYSICS

Speech by Professor SVEN JOHANSSON of the Royal Academy of Sciences. Translation from the Swedish text

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen,

The problem of the basic structure of matter has long interested man but it was not until the time of the Greek philosophers that the problem took on a scientific character. These ideas reached their culmination in Democritos' theory which postulated that atoms were the building blocks of matter. All this was, however, mere speculation, and it was first the early science and technology of Western Europe which made it possible to tackle the problem experimentally.

The first major breakthrough came with the invention of the microscope. The significance of the microscope in the fields of, for example, biology and medicine is well known, but it did not provide a means of studying the basic nature of matter. The reason is that there is a limit to the amount of detail one can see in a microscope. This is connected with the wave nature of light. In the same way as ocean waves are not affected, to any great degree, by small objects, but only by larger ones, for example a breakwater, light will not produce a picture of an object that is too small. The limit is set by the wavelength of light which is about 0.0005 mm. We know that an atom is 1000 times smaller. It is clear, therefore, that something radically new was needed in order to be able to see an atom.

This new development was the electron microscope. The electron microscope is based on the principle that a short coil of a suitable construction, carrying an electric current, can deflect electrons in the same way that a lens deflects light. A coil can therefore give an enlarged image of an object that is irradiated with electrons. The image can be registered on a fluorescent screen or a photographic film. In the same way that lenses can be combined to form a microscope, it was found that an electron microscope could be constructed of coils. As the electrons used in an electron microscope have a much shorter wavelength than light, it is thus possible to reach down to much finer details. Several scientists, among them Hans Busch, Max Knoll, and Bodo von Borries, contributed to the development of the instrument, but Ernst Ruska deserves to be placed foremost. He built in 1933 the first electron microscope with a performance significantly better than that of an ordinary light microscope. Developments since then have led to better and better instruments. The importance, in many areas of research, of the invention of the electron microscope should, by now, be well known.

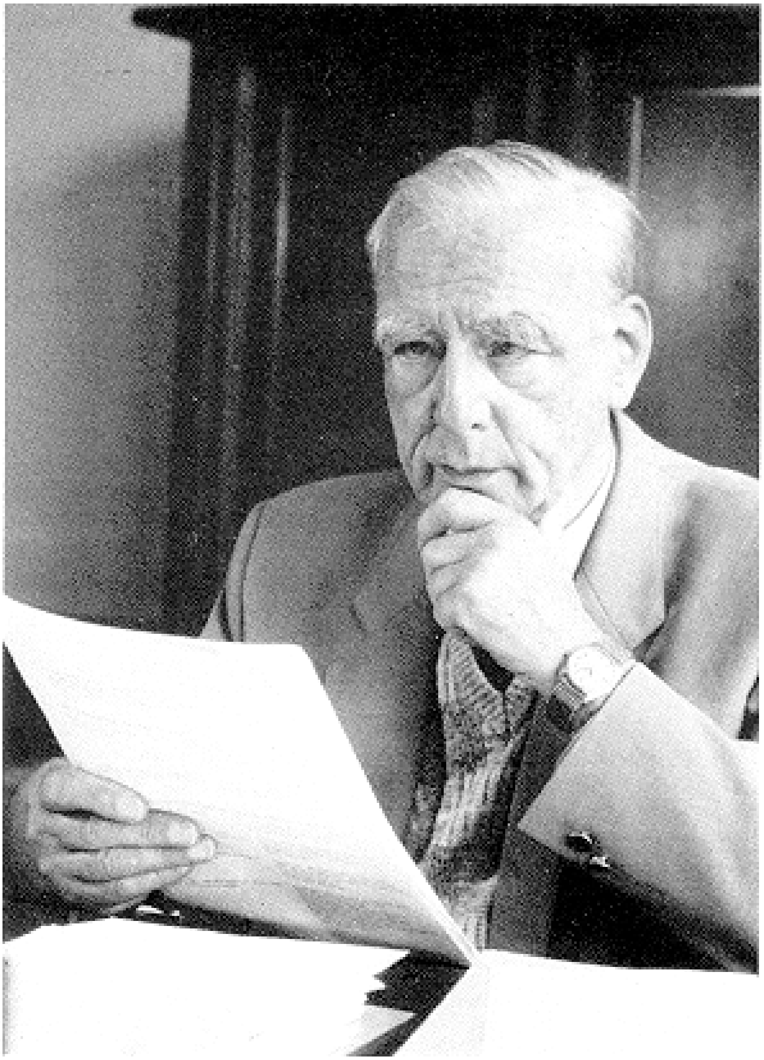
The microscope can be regarded as an extension of the human eye. But sight is not the only sense we use to orientate us in our surroundings, another is feeling. With modern technology it is possible to construct equipment that is based on the principle of feeling, using, for example, a sort of mechanical linger. The "finger" may be a very fine needle which is moved across the surface of the structure to be investigated. By registering the needle's movements in the vertical direction as it traverses the surface, a sort of topographical map is obtained, which, in principle, is equivalent to the image obtained in a electron microscope. It is clear that this is a rather coarse method of microscopical investigation and no one had expected any revolutionary developments in this field. However, two basic improvements led to a breakthrough. The most important of these was that a method for keeping the tip of the needle at a very small and exact constant distance from the surface was developed, thus eliminating the mechanical contact between the needle and the surface, which was a limiting factor. This was achieved using the so-called tunnelling effect. This involves applying a potential between the needle tip and the surface so that an electric current flows between the needle and the surface without actually touching them, provided that the tip of the needle and the surface are close enough together. The magnitude of the current is strongly dependent on the distance, and can therefore be used to keep the needle a certain distance above the surface with the aid of a servo mechanism, typically 2-3 atomic diameters. It was also decisive that it turned out to be possible to produce extremely fine needles so that the tip consists of only a few atoms. It is clear that if such a fine tip is moved across a surface at a height of a few atomic diameters the finest atomic details in the surface structure can be registered. It is as if one were feeling the surface with an infinitely fine finger. A crystal surface which appears completely flat in a microscope is seen with this instrument to be a plain on which **atoms** rise like hills in a regular pattern.

Attempts by Russell Young and co-workers to realize these ideas revealed enormous experimental difficulties. The scientists who finally mastered these difficulties were Gerd Binnig and Heinrich Rohrer. Here it was a question of moving the needle over the surface of the sample and registering its vertical position, with great precision and without disturbing vibrations. The data obtained arc then printed out, in the form of a topographic map of the surface, by a computer. The investigation may be concerned with a crystal surface, whose structure is of interest in microelectronic applications. Another example is the investigation of the adsorption of atoms on a surface. It has also been found to be possible to study organic structures, for example, DNA molecules and viruses. This is just the beginning of an extremely promising and fascinating development. The old dream from antiquity of a visible image of the atomic structure of matter is beginning to look like a realistic possibility, thanks to progress in modern microscopy.

Professor Ruska, Dr Binnig, Dr Rohrer!

In Ihrer bahnbrechenden Arbeit haben Sie den Grund fir die entscheidenden Entwicklungen moderner Mikroskopie gelegt. Es ist jetzt möglich, die kleinsten Einzelheiten der Struktur von Materie zu erkennen. Dies ist von grösster Bedeutung - nicht nur in der Physik, sondern such in vielen anderen Bereichen der Wissenschaft.

Es gereicht mir zur Ehre und Freude, Ihnen die herzlichsten Glöckwünsche der Königlich Schwedischen Akademie der Wissenschaften zu iibermitteln. Darf ich Sic nun bitten vorzutreten, urn Ihren Preis aus der Hand Seiner Majestät des Königs entgegenzunehmen.



Dr. Ernst Aucke

ERNST RUSKA

I was born on 25 December 1906 in Heidelberg as the fifth of seven children of Professor Julius Ruska and his wife Elisbeth (née Merx). After graduating from grammar school in Heidelberg I studied electronics at the Technical College in Munich, studies which I began in the autumn of 1925 and continued two years later in Berlin. I received my practical training from Brown-Boveri & Co in Mannheim and Siemens & Halske Ltd in Berlin. Whilst still a student at the Technical College in Berlin I began my involvement with high voltage and vacuum technology at the Institute of High Voltage, whose director was Professor Adolf Matthias. Under the direct tutelage of Dr Max Knoll and together with other doctoral students I worked on the development of a high performance cathode ray oscilloscope. On the one hand my interest lay principally in the development of materials for the building of vacuum instruments according to the principles of construction; on the other it lay in continuing theoretical lectures and practical experiments in the optical behaviour of electron rays.

My first completed scientific work (1928-9) was concerned with the mathematical and experimental proof of Busch's theory of the effect of the magnetic field of a coil of wire through which an electric current is passed and which is then used as an electron lens. During the course of this work I recognised that the focal length of the waves could be shortened by use of an iron cap. From this discovery the polschuh lens was developed, a lens which has been used since then in all magnetic high-resolution electron microscopes. Further work, conducted together with Dr Knoll, led to the first construction of an electron microscope in 1931. With this instrument two of the most important processes for image reproduction were introduced - the principles of emission and radiation. In 1933 I was able to put into use an electron microscope, built by myself, that for the first time gave better definition than a light microscope. In my Doctoral thesis of 1934 and for my university teaching thesis (1944) both at the Technical College in Berlin, I investigated the properties of electron lenses with short focal lengths.

Since the further technical development of electron microscopes could not be the task of a college institute -whose resources would have been far overstretched-1 went to work in industry in the field of electron optics. From 1933 to 1937 I was with Fernseh Ltd in Berlin-Zehlendorfand was responsible for the development of television receivers and transmitters, as well as photoelectric cells with secondary amplification. Convinced of the great practical importance of electron microscopy for pure and applied research I attempted during this time to continue the development of high-resolution electron microscopes with larger materials, this time working with Dr Bodo von Borries. This work was made possible in 1936-7 by Siemens & Halske. In Berlin-Spandau in 1937 we set up the Laboratory for Electron Optics and developed there until 1939 the first customised electron microscopes (the 'Siemens Super Microscope'). Parallel to the development of this instrument my brother, Dr Med. Helmut Ruska, and his colleagues worked on its application, particularly in the medical and biological fields. In order to promote its usage in different scientific areas as quickly as possible we suggested to Siemens that they set up a visiting institute for research work to be carried out using electron microscopy. This institute was founded in 1940. From this institute, in which we worked together with both German and foreign scientists, around 200 scientific paperswere published before the end of 1944. My task consisted in the development and production of the electron microscope, such that by the beginning of 1945 around 35 institutions were equipped with one.

In the years following 1945 I, together with a majority of new colleagues, reconstituted the Institute of Electron Optics in Berlin-Siemensstadt, which had been disbanded due to bombing, so that by 1949 electron microscopes were again being built. This new period of development led in 1954 to `Elmiskop l', which since then has been used in over 1200 institutions the world over. At the same time I sought the further physical development of the electron microscope by working at other scientific institutions. Thus from August 1947 to December 1948 I worked at the German Academy of Sciences in Berlin-Buch in the Faculty of Medicine and Biology, then from January 1949 as Head of Department at what is today the Fritz Haber Institute of the Max Planck Society in Berlin-Dahlem. Here on 27 June 1957 I was made Director of the Institute for Electron Microscopy, after I had given up my position with Siemens in 1955. I retired on 31 December 1974.

From 1949 until 1971 I held lectures on the basic principles of electron optics and electron microscopy at both the Free University and the Technical University of Berlin. My publications in the area of electron optics and electron microscopy include several contributions to books and over 100 original scientific papers.

(added by *the editor)* : Ernst Ruska died on May 25, 1988.

THE DEVELOPMENT OF THE ELECTRON MICROSCOPE AND OF ELECTRON MICROSCOPY

Nobel lecture, December 8, 1986

by

ERNST RUSKA

Max-Eyth-Strasse 20, D-1000 BERLIN 33

A. Parents'house, family

A month ago, the Nobel Foundation sent me its yearbook of 1985. From it I learnt that many Nobel lectures are downright scientific lectures, interspersed with curves, synoptic tables and quotations. I am somewhat reluctant to give here such a lecture on something that can be looked up in any modern schoolbook on physics. I will therefore not so much report here on physical and technical details and their connections but rather on the human experiences-some joyful events and many disappointments which had not been spared me and my colleagues on our way to the final breakthrough. This is not meant to be a complaint though; I rather feel that such experiences of scientists in quest of new approaches are absolutely understandable, or even normal.

In such a representation I must, of course, consider the influence of my environment, in particular of my family. There have already been some scientists in my family: My father, Julius Ruska, was a historian of sciences in Heidelberg and Berlin; my uncle, Max Wolf, astronomer in Heidelberg; his assistant, a former pupil of my father and my godfather, August Kopff, Director of the Institute for astronomical calculation of the former Friedrich-Wilhelm University in Berlin. A cousin of my mother, Alfred Hoche, was Professor for Psychiatry in Freiburg/Breisgau; my grandfather from my mother's side, Adalbert Merx, theologian in Gießen and Heidelberg.

My parents lived in Heidelberg and had seven children. I was the fifth, my brother Helmut the sixth. To him I had particularly close and friendly relations as long as I can remember. Early, optical instruments made a strong impression on us. Several times Uncle Max had shown us the telescopes at the observatory on the Königstuhl near Heidelberg headed by him. With the light microscope as well we soon had impressive, yet contradictory, relations. In the second floor of our house, my father had two study rooms connected by a broad sliding door which usually was open. One room he used for his scientific historical studies relating to classical philology, the other for his scientific interests, in particular mineralogy, botany and zoology. When our games with neighbours' kids in front of the house became too noisy, he would knock at the window panes. This usually only having a brief effect, he soon knocked a second time, this time considerably louder. At the third knock, Helmut and I had to come to his room and sit still on a low wooden stool, dos à dos, up to one hour at 2 m distance from his desk. While doing so we would see on a table in the other room the pretty yellowish wooden box that housed my father's big Zeiss microscope, which we were strictly forbidden to touch. He sometimes demonstrated to us interesting objects under the microscope, it is true; for good reasons, however, he feared that childrens' hands would damage the objective or the specimen by clumsy manipulation of the coarse and line drive. Thus, our first relation to the value of microscopy was not solely positive.

B. School, vocational choice

Much more positive was, several years later, the excellent biology instruction my brother had through his teacher Adolf Leiber and the very thorough teaching I received through my teacher Karl Reinig. To my great pleasure I recently read an impressive report on Reinig's personality in the Memoirs of a two-years-older student at my school, the later theoretical physicist Walter Elsasser. Even today I remember the profound impression Reinig's comments made upon me when he explained that the movement of electrons in an electrostatic field followed the same laws as the movement of inert mass in gravitational fields. He even tried to explain to us the limitation of microscopical resolution due to the wavelength of light. I certainly did not clearly understand all this then, because soon after that on one of our many walks through the woods around Heidelberg I had a long discussion on that subject with my brother Helmut, who already showed an inclination to medicine, and my classmate Karl Deißler, who later studied medicine as well.

In our College (Humanistisches Gymnasium), we had up to 17 hours of Latin, Greek and French per week. In contrast to my father, who was extremely gifted for languages, I produced only very poor results in this field. My father, at that time teacher at the same school, daily learnt about my minus efforts from his colleagues and blamed me for being too lazy, so that I had some sorrowful school years. My Greek teacher, a fellow student of my father, had a more realistic view of things: He gave me for my confirmation the book "Hinter Pflug und Schraubstock" (Behind plow and vise) by the Swabian "poet" engineer Max Eyth (1836- 1906). I had always been fascinated by technical progress; in particular I was later interested in the development of aeronautics, the construction of airships and air planes. The impressive book of Max Eyth definitely prompted me to study engineering. My father, having studied sciences at the universities of Straßburg, Berlin and Heidelberg, obviously regarded study at a Technical High School as not being adequate and offered me one physics semester at a university. I had, however, the strong feeling that engineering was more to my liking and refused.

C. The cathode-ray oscillograph and the short coil

After I had studied two years electrotechnical engineering in Munich, my father received a call to become head of a newly founded Institute for the

History of Sciences in Berlin in 1927. Thus, after my pre-examination in Munich I came to Berlin for the second half of my studies. Here I specialized in high-voltage techniques and electrical plants and heard, among others, the lectures of Professor Adolf Matthias. At the end of the summer term in 1928 he told us about his plan of setting up a small group of people to develop from the Braun tube an efficient cathode-ray oscillograph for the measurement of very fast electrical processes in power stations and on open-air high-voltage transmission lines. Perhaps with the memory of my physics school lesson in the back of my head, I immediately volunteered for this task and became the youngest collaborator of the group, which was headed by Dr. Ing. Max Knoll. My first attempts with experimental work had been made in the practical physics course at the Technical High School in Munich under Professor Jonathan Zenneck, and now in the group of Max Knoll. As a newcomer I was first entrusted with some vacuum-technical problems which were important to all of us. Through the personality of Max Knoll, there was a companionable relationship in the group, and at our communal afternoon coffee with him the scientific day-to-day-problems of each member of the group were openly discussed. As I did not dislike calculations, and our common aim was the development of cathode-ray oscillographs for a desired measuring capability, I wanted to devise a suitable method of dimensioning such cathode-ray oscillographs in my “Studienarbeit”-a prerequisite for being allowed to proceed to the Diploma examination.

The most important parameters for accuracy of measurement and writing speed af cathode-ray oscillographs are the diameter of the writing spot and its energy density. To produce small and bright writing spots, the electron beams emerging divergently from the cathode had to be concentrated in a small writing spot on the fluorescent screen of the cathode-ray oscillograph. For this, already Rankin in 1905 [1] used a short dc-fed coil, as had been used by earlier experimentalists with electron beams (formerly called “glow” or “cathode rays”). Even before that, Hittorf (1869) and Birkeland (1896) used the rotationally symmetric field lying in front of a cylindrical magnet pole for focussing cathode rays. A more precise idea of the effect of the axially symmetric, i.e. inhomogeneous magnet field of such poles or coils on the electron bundle alongside of their axes had long been unclear.

Therefore, Hans Busch [3] at Jena calculated the electron trajectories in such an electron ray bundle and found that the magnetic field of the short coil has the same effect on the electron bundle as has the convex glass lens with a defined focal length on a light bundle. The focal length of this “magnetic electron lens” can be changed continuously by means of the coil current. Busch wanted to check experimentally his theory but for reasons of time he could not carry out new experiments. He made use of the experimental results he had already obtained sixteen years previously in Gottingen. These were, however, in extremely unsatisfactory agreement with the theory. Perhaps this was the reason that Busch did not draw at least the practical conclusion from his lens theory to image some object with such a coil.

In order to account more precisely for the properties of the writing spot of a cathode-ray oscillograph produced by the short coil, I checked Busch's lens theory with a simple experimental arrangement under better, yet still inadequate, experimental conditions (Fig. 1) and thereby found a better but still not entirely satisfactory agreement of the imaging scale with Busch's theoretical

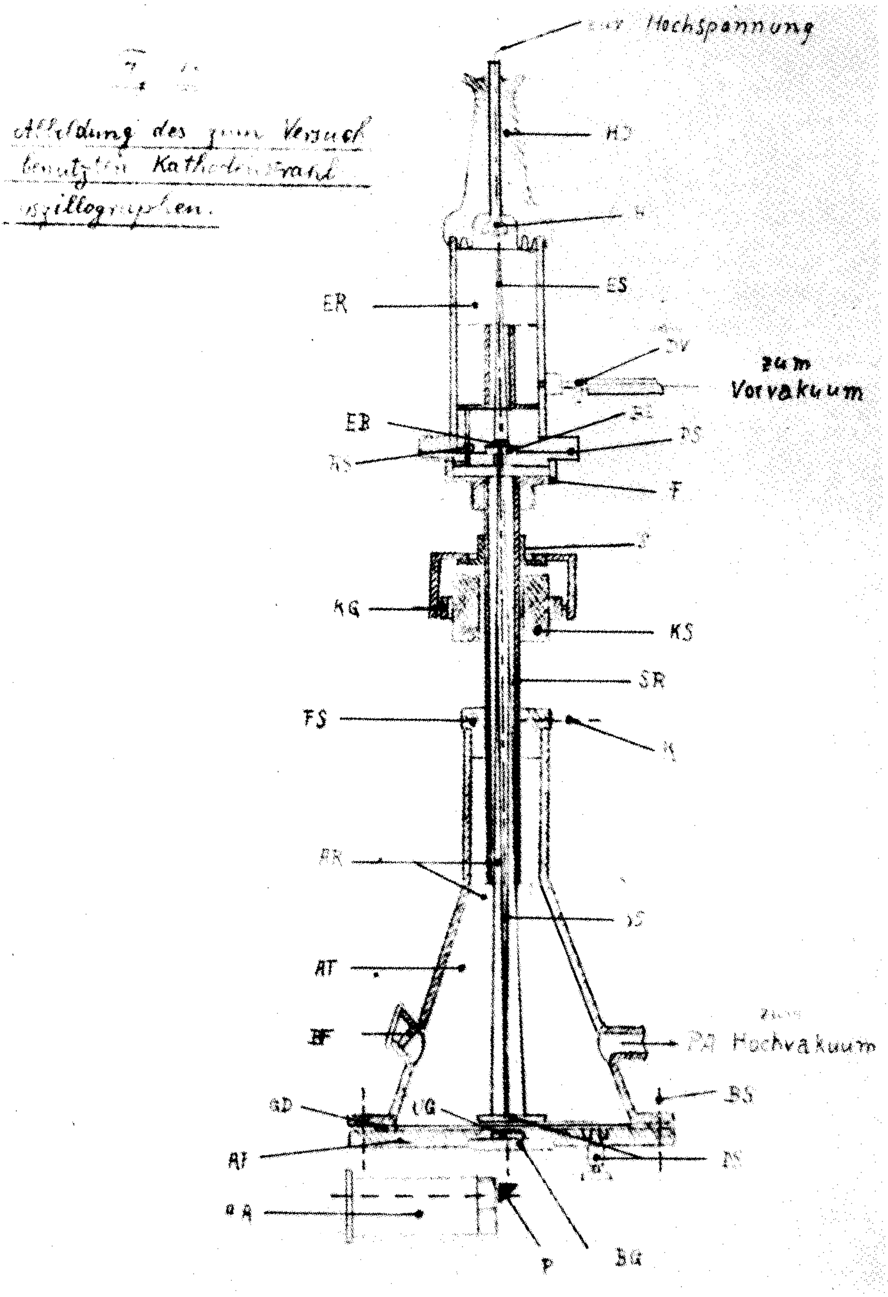


Fig. 1: Sketch by the author (1929) of the cathode ray tube for testing the imaging properties of the non-uniform magnetic field of a short coil [4. 5].

expectation. The main reason was that I had used a coil of the dimensions of Busch's coil whose field distribution along the axis was much too wide. My Studienarbeit [4], submitted to the Faculty for Electrotechnical Engineering in 1929, contained numerous sharp images with different magnifications of an electron-irradiated anode aperture of 0.3 mm diameter which had been taken by means of the short coil ("magnetic electron lens")-i. e. the first recorded electron-optical images.

Busch's equation for the focal length of the magnetic field of a short coil implied that a desired focal length could be produced by the fewer ampere turns the more the coil field was limited to a short region alongside the axis, because in that case the field maximum is increased. It was therefore logical for me as a prospective electrotechnical engineer to suitably envelop the coil with an iron coating, with a ring-shaped gap in the inner tube. Measurements at such a coil immediately showed that the same focal length had been reached with markedly fewer ampere turns [4,5]. Vice versa, in this manner a shorter focal length can, of course, also be obtained by an equal number of ampere turns.

D. Why I pursued the magnetic electron lens for the electron microscope

In my Diploma Thesis (1930) I was to search for an electrostatic replacement for the magnetic concentration of the divergent electron ray bundle, which would probably be easier and cheaper. To this end, Knoll suggested experimental investigation of an arrangement of hole electrodes with different electrical potential for which he had taken out a patent a year before [6]. We discussed the shape of the electric field between these electrodes, and I suggested that because of the mirror-like symmetry of the electrostatic field of the electrodes on either side of the lens centre, a concentrating effect of the curved equipotential planes in the hole area could not take place. I only had the field geometry in mind then. But this conclusion was wrong. I overlooked that as a consequence of the considerably varying electron velocity on passage through such a field arrangement, a concentration of the divergent electron bundle must, in fact, occur. Knoll did not notice this error either. Therefore I pursued another approach in my Diploma Thesis [7]. I made the electron bundle pass a bored-out spherical condenser with fine-meshed spherically shaped grids fixed over each end of the bore. With this arrangement I obtained laterally inverted images in the correct imaging scale. Somewhat later I found a solution which was unfortunately only theoretically correct. In analogy to the refraction of the light rays on their passage through the optical lens at their surfaces ("Grenzflächen), I wanted to use, for the electrical lens, the potential steps at corresponding surfaces, which are shaped like glasses lenses [8]. Thus, the energy of the electron beams is temporarily changed-just like light beams on passage through optical lenses. For the realization of this idea, on each side of the lens two closely neighboured fine-meshed grids of the shape of optical lenses are required which must be kept on electrical potentials different from each other. First attempts confirmed the rightness of this idea, but at the same time also the practical inaptness of such grid lenses because of the too-strong absorption of the electron beam at the four grids and due to the field distribution by the wires.

As a consequence of my false reasoning and the experimental disappointment I decided to continue with the magnetic lens. I only report this in so much detail to show that occasionally it can be more a matter of luck than of superior intellectual vigor to find a better-or perhaps the only acceptable way. The approach of the transmission electron microscope with electron lenses of electrostatic hole electrodes was later pursued by outstanding experimentalists in other places and led to considerable initial success. It had, however, to be abandoned because the electrostatic lens was for physical reasons inferior to the magnetic electron lens.

E. The invention of the electron microscope

After obtaining my Degree (early 1931), the economic situation had become very difficult in Germany and it seemed not possible to find a satisfactory position at a University or in industry. Therefore I was glad that I could at least continue my unpaid position as doctorand in the high-voltage institute. After having shown in my Studienarbeit of 1929 that sharp and magnified images of electron-irradiated hole apertures could be obtained with the short coil, I was now interested in finding out if such images-as in light optics-could be further magnified by arranging a second imaging stage behind the first stage. Such an apparatus with two short coils was easily put together (Fig. 2) and in April 1931 I obtained the definite proof that it was possible (Fig. 3). This apparatus is justifiably regarded today as the first electron microscope even though its total magnification of 3.6 x 4.8 = 14.4 was extremely modest.

The first proof had thus been given that-apart from light and glass lenses-images of irradiated specimens could be obtained also by electron beams and magnetic fields, and this in even more than one imaging stage. But what was the use of such images if even grids of platinum or molybdenum were burnt to cinders at the irradiation level needed for a magnification of only 17.4 x. Not wishing to be accused of showmanship, Max Knoll and I agreed to avoid the term *electron microscope* in the lecture Knoll gave in June 1931 on the progress in the construction of cathode ray oscillographs where he also, for the first time, described in detail my electron-optical investigations [9, 10]. But, of course, our thoughts were circling around a more efficient microscope. The resolution limit of the light microscope due to the length of the light wave which had been recognized 50 years before by Ernst Abbe and others could, because of lack of light, not be important at such magnifications. Knoll and I simply hoped for extremely low dimensions of the electrons. As engineers we did not know yet the thesis of the “material wave” of the French physicist de Broglie [11] that had been put forward several years earlier (1925). Even physicists only reluctantly accepted this new thesis. When I first heard of it in summer 1931, I was very much disappointed that now even at the electron microscope the resolution should be limited again by a wavelength (of the "Materiestrahlung"). I was immediately heartened, though, when with the aid of the de Broglie equation I became satisfied that these waves must be around five orders of

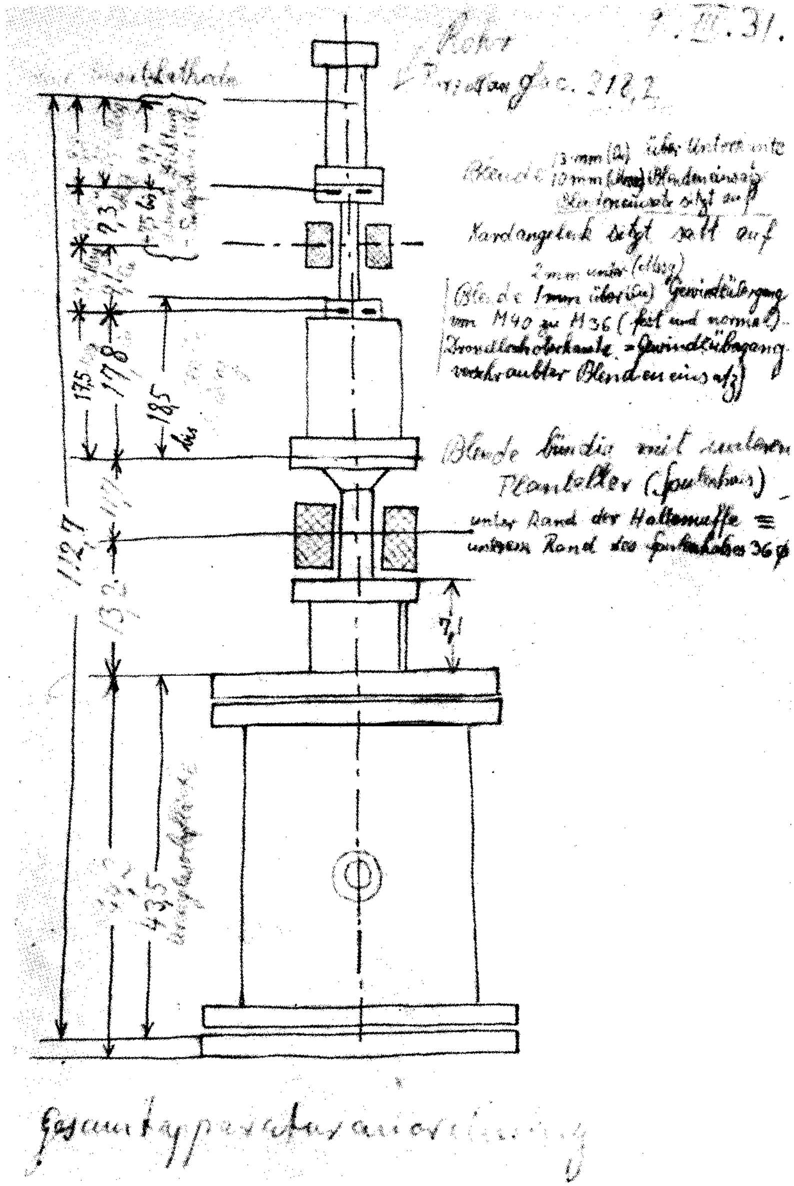


Fig. 2: Sketch by the author (9 March 1931) of the cathode ray tube for testing one-stage and two-stage electron-optical imaging by means of two magnetic electron lenses (electron microscope) [8].

magnitude shorter in length than light waves. Thus, there was no reason to abandon the aim of electron microscopy surpassing the resolution of light microscopy.

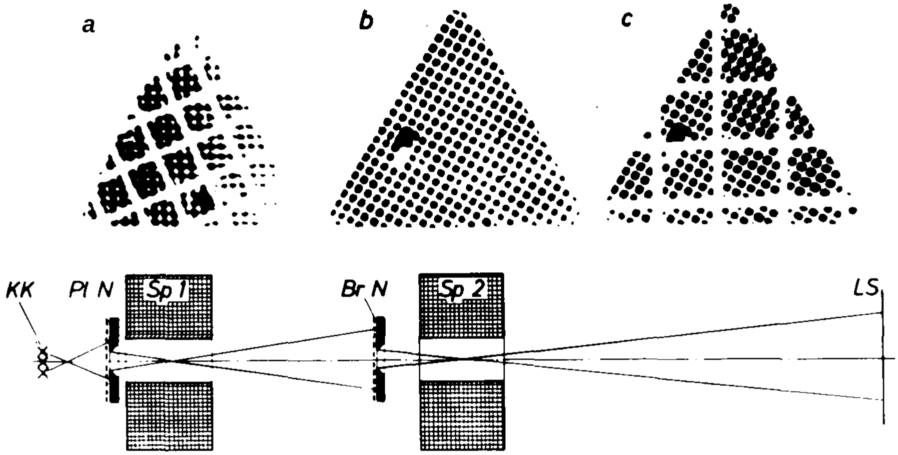


Fig. 3: First experimental proof (7 April 1931) that speciemens (aperture grids) irradiated by electrons can be imaged in magnified form not only in one but also in more than one stage by means of (magnetic) electron lenses.
(U = 50 kV). [8].
a) one-stage image of the platinum grid in front of coil 1 by coil 1; M = 13 x
b) one-stage image of the bronze grid in front of coil 2 by coil 2; M = 4.8 x
c) two-stage image of the platinum grid in front of coil 1 by coil 2; M = 17.4 x together with the one-stage image of the bronze grid in front of coil 2 by coil 2; M = 4.8 x
kk Cold cathode; Pt N Platinum grid; Sp 1 coil 1;
Br N Bronze grid; Sp 2 coil 2; LS Fluorescent screen

In 1932 Knoll and I dared to make a prognosis of the resolution limit of the electron microscope [12]. Assuming that the equation for the resolution limit of the light microscope is valid also for the material wave of the electrons, we replaced the wave length of the light by the wave length of electrons at an accelerating voltage of 75 kV and inserted into the Abbe relation the imaging aperture of 2×10^{-2} rad which is what we had used previously. This imaging aperture is still used today. Thereby, that early we came up with a resolution limit of 2.2 Å = 2.2×10^{-10}m, a value that was in fact obtained 40 years later.

Of course, at that time our approach was not taken seriously by most of the experts. They rather regarded it as a pipe-dream. I myself felt that it would be very hard to overcome the efforts still needed-mainly the problem of specimen heating. In April 1932, M. Knoll had taken up a position with Telefunken (Berlin) involving developmental work in the field of television.

In contrast to many biologists and medical scientists, my brother Helmut, who had almost completed his medical studies, believed in considerable progress for these disciplines should we be successful. With his confidence in a successful outcome he encouraged me to overcome the expected difficulties. In a next step I had to show that it was possible to obtain sufficiently high magnifications to prove a better-than-light-microscope resolution. To this effect a coil shape had to be developed whose magnetic field was compressed to a length that small of the coil axis to allow short focal lengths as are needed for

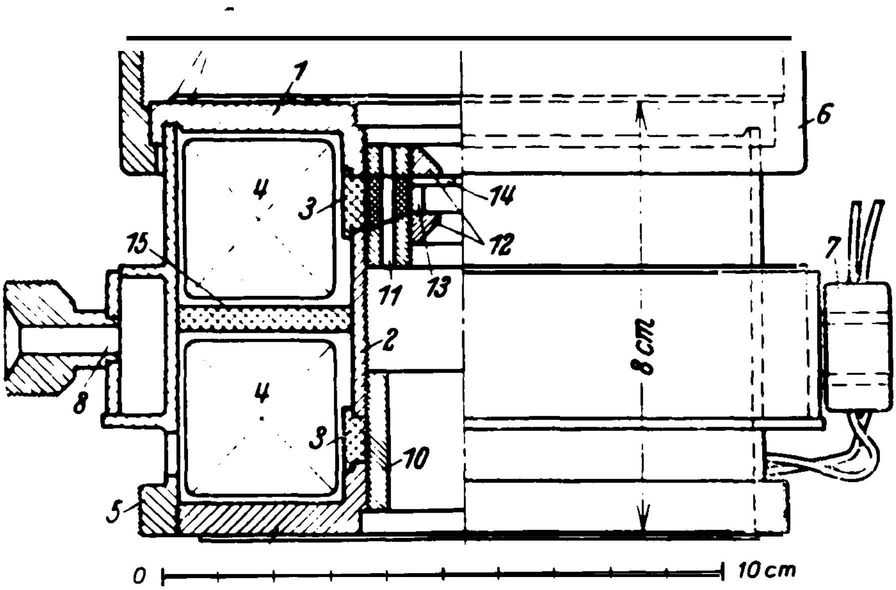


Fig. 4: Cross-section of the first polepiece lens [4, 15].

highly magnified images in not too great a distance behind the coil. The technical solution for this I had already given in my Studienarbeit of 1929 with the iron-clad coil. In 1932 I applied-together with my friend and co-doctorand Bodo v. Borries-for a patent on the optimization of this solution[l3], the "Polschuhlinse", which is used in all magnetic electron microscopes today. Its realization and the measuring of the focal lengths which could be verified with it were subject of my thesis [14]. It was completed in August 1933, and in my measurements I obtained focal lengths of 3 mm for electron rays of 75 kV acceleration (Fig. 4). Of course, now with these lenses I immediately wanted to design a second electron microscope with much higher resolving power. To carry out this task I obtained by the good offices of Max v. Laue for the second half year of 1933 a stipend of Reichsmark 100 per month from the Notgemeinschaft der Deutschen Wissenschaft to defray running costs and personal expenses. Since I had completed the new instrument by the end of November (Fig. 5), I felt I ought to return my payment for December. To my great joy, however, I was allowed to keep the money "as an exception". Nevertheless, this certainly was the cheapest electron microscope ever paid for by a German organization for the promotion of science.

For reasons explained in the beginning of the next chapter, I accepted a position in industry on 1 December 1933. Therefore I could only make a few images with this instrument which magnified 12000 × [15], but I noticed a decisive fact which gave me hope for the future: Even very thin specimens yielded sufficient contrast, yet no longer by absorption but solely by diffraction of the electrons, whereby-as is known-the specimens are heated up considerably less.

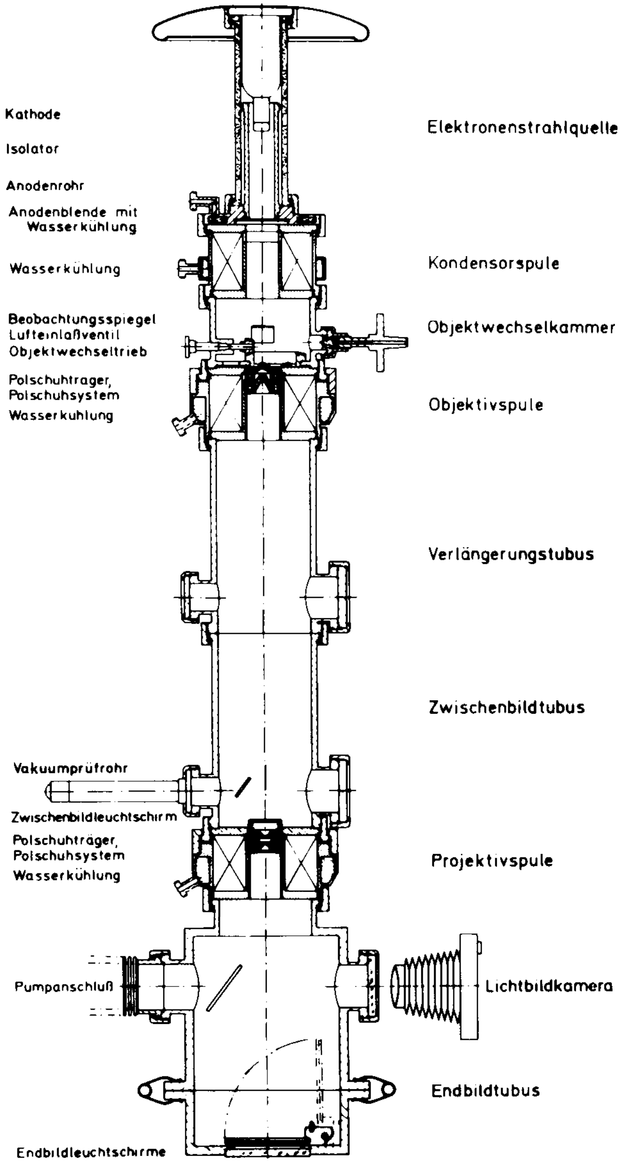


Fig. 5: First (two-stage) electron microscope magnifying higher than the light microscope. Cross-section of the microscope column (Re-drawn 1976) [15].

F: How the industrial production of electron microscopes *came to be*

I also realized, however, that the further development of a practically useful instrument with better resolution would require a longer period of time and enormous costs. In view of the results achieved there was little hope of obtaining financial support from any side for the time being. I was prepared for a longer dry spell and decided to approach the goal of a commercial instrument later, together with Bodo v. Borries and my brother Helmut. Therefore, I

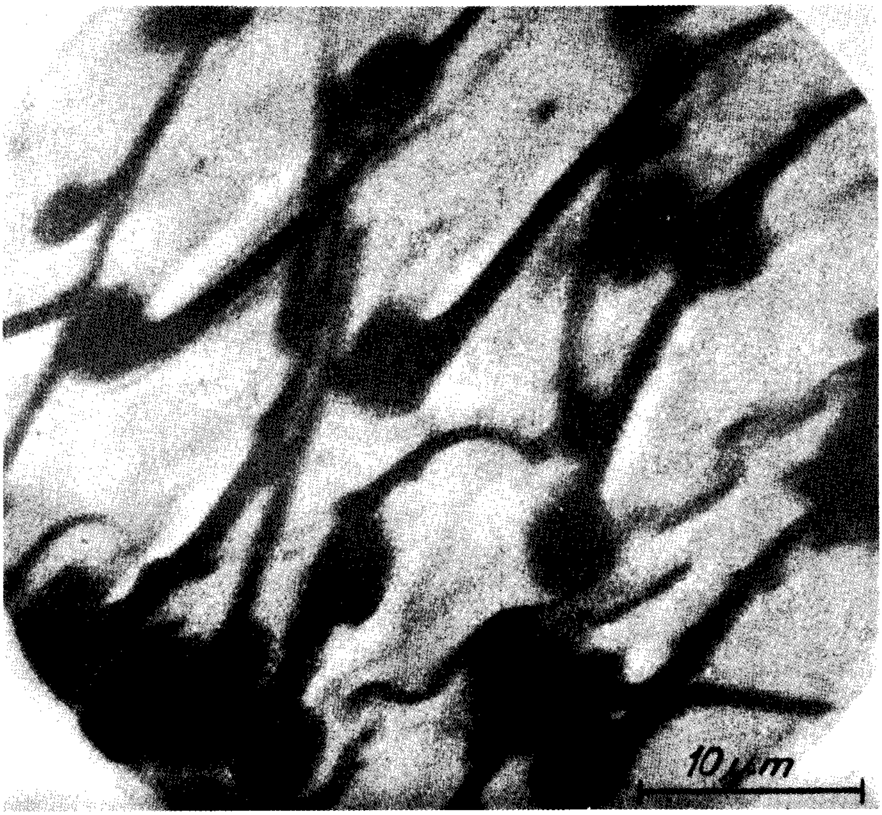


Fig. 6: Wing surface of the house fly.
(First internal photography, U = 60 kV, M_{el} = 2200)
(Driest, E., and Müller, H.O.: Z. Wiss. Mikroskopie 52, 53-57 (1935)

accepted a position with the Fernseh AG in Berlin-Zehlendorf where I was engaged in the development of Braun tubes for image pick-up and display tubes. In order to better coordinate our efforts to obtain financial support for the production of commercial electron microscopes, I convinced Bodo v. Borries to give up his position at the Rheinisch-Westfalische Elektrizitatswerke at Essen and return to Berlin. Here, he found a position at Siemens-Schuckert in 1934. We approached many governmental and industrial research facilities for financial help.

During this period, first electron micrographs appeared of biological specimens. Heinz Otto Müller (student in electrotechnical engineering) and Friedrich Krause (medical student) worked at the instrument I had built in 1933, and they published increasingly better results (Figs. 6 to 9). Unfortunately these two very gifted young scientists did not survive the II. World War.

At Brussels Ladislaus Marton had built his first horizontal microscope and obtained relatively low magnifications of biological specimens [17]. In 1936 he built a second instrument, this time with a vertical column [18].

Fig. 7: Diatoms ***Amphipleura pellucida.***
(U = 53 kV, M_{el} = 3500, δ'' = 130 nm)
(F. Krause in: Busch, H., and Brüche, E.: Beiträgr zur Elektronenoptik, 55-61, Verl. Joh. Ambrosius Barth, Leipzig 1937)

Fig. 8: Bacteria (culture infusion), fixed with formalin and embedded in a supporting film stained with a heavy metal salt
(U = 73.5 kV, M_{el} = 2000)
(Krause, F.: Naturwissenschaften 25,817-825 (1937)).

In spite of these more recent publications, it took us three years to be successful in our quest for financial support through the professional assessment of Helmut Ruska's former clinical teacher, Professor Dr. Richard Siebeck, Director of the I. Medical Clinic of the Berlin Charité. I quote two paragraphs of his assessment of 2 October 1936 [19]:

"If these things were to be realised it hardly needs to be emphasised that the advances in the field of research into the causes of disease would be of immediate practical interest to the doctor. It would deeply affect real problems concerned to a large extent with diseases of growing clinical significance and thus of great importance for public health.

Should the possibilities of microscopical resolution exceed the assumed values by a factor of a hundred, the scientific consequences would be incalculable. What seems attainable now, I consider to be so important, and success seems to me so close, that I am ready and willing to advise on medical research work and to collaborate by making available the resources of my Institute".

Fig. 9.: Iron Whisker
(U = 79 kW, M_{el} = 3100)
(Beischer, D., and Krause, F.: Naturwissenschaften 25, 825-829 (1937)).

This expertise impressed Siemens in Berlin and Carl Zeiss in Jena, and they were both ready to further the development of industrial electron microscopes. We suggested the setting up of a common development facility in order to make use of the electrotechnical expertise of Siemens and the know-how in precision engineering of Zeiß, but unfortunately the suggestion was refused and so we decided in favour of Siemens. As first collaborators we secured Heinz Otto Müller for the practical development and Walter Glaser from Prag as theorist. We started in 1937, and in 1938 we had completed two prototypes with condenser and polepieces for objective and projective as well as airlocks for specimens and photoplates. The maximum magnification was 30000 × [20]. One of these instruments was immediately used for first biological investigations by Helmut Ruska and several medical collaborators. (H. Ruska was released from Professor Siebeck for our work at Siemens.) Unfortunately, for reasons of time I cannot give here a survey of this fruitful publication period.

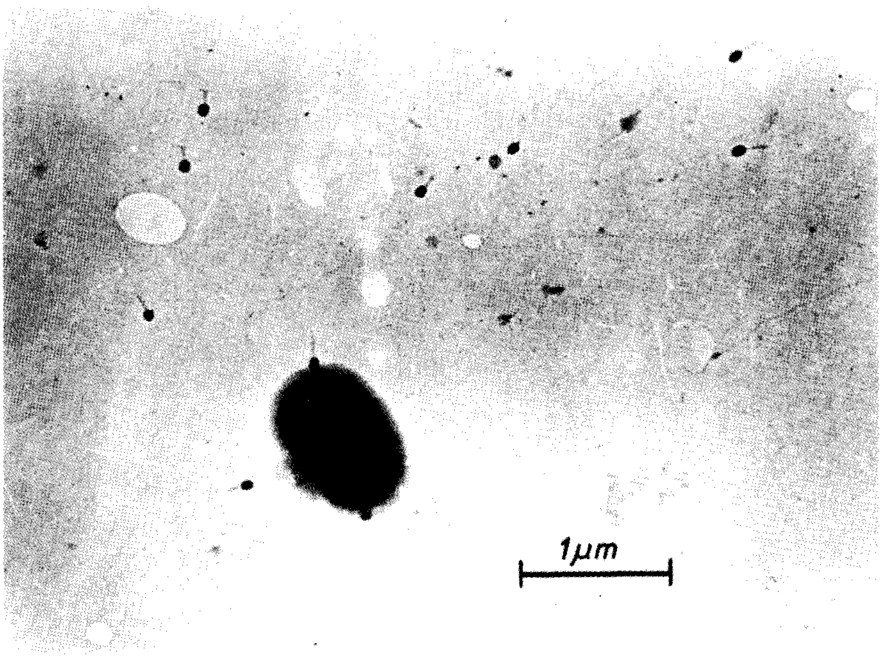


Fig. *10:* Bacteriophages. (Ruska, H.: Naturwissenschaften 29, 367-368 (1941) and Arch. Ges. Virusforsch 2, 345-387 (1942).

In 1940, upon our proposal Siemens set up a guest laboratory, headed by Helmut Ruska, with four electron microscopes for visiting scientists. Helmut Ruska could show first images of bacteriophages in 1940. An image taken somewhat later (Fig. 10) clearly shows the shape of these tiny hostile bacteria. This laboratory was destroyed during an air raid in the autumn of 1944.

Very gradually now interest in electron microscopy was growing. A first sales success for Siemens has been achieved in 1938 when the chemical industry which was represented largely by IG Farbenindustrie placed orders for an instrument in each of their works in Hoechst, Leverkusen, Bitterfield and Wolfen. The instrument was only planned at the time, however not yet built or even tested. By the end of 1939 the first serially produced Siemens instrument [21] had been delivered to Hoechst (Fig. 11). The instrument No 26 was, by the way, delivered to Professor Arne Tiselius in Uppsala in autumn 1943. By February 1945 more than 30 electron microscopes had been built in Berlin and delivered. Thus, now also independent representatives of various medical and biological disciplines could form their own opinions about the future prospects of electron microscopy. The choice of specimens was still limited though, since sufficiently thin sections were not yet available. The end of the war terminated the close cooperation with my brother and B. v. Borries.

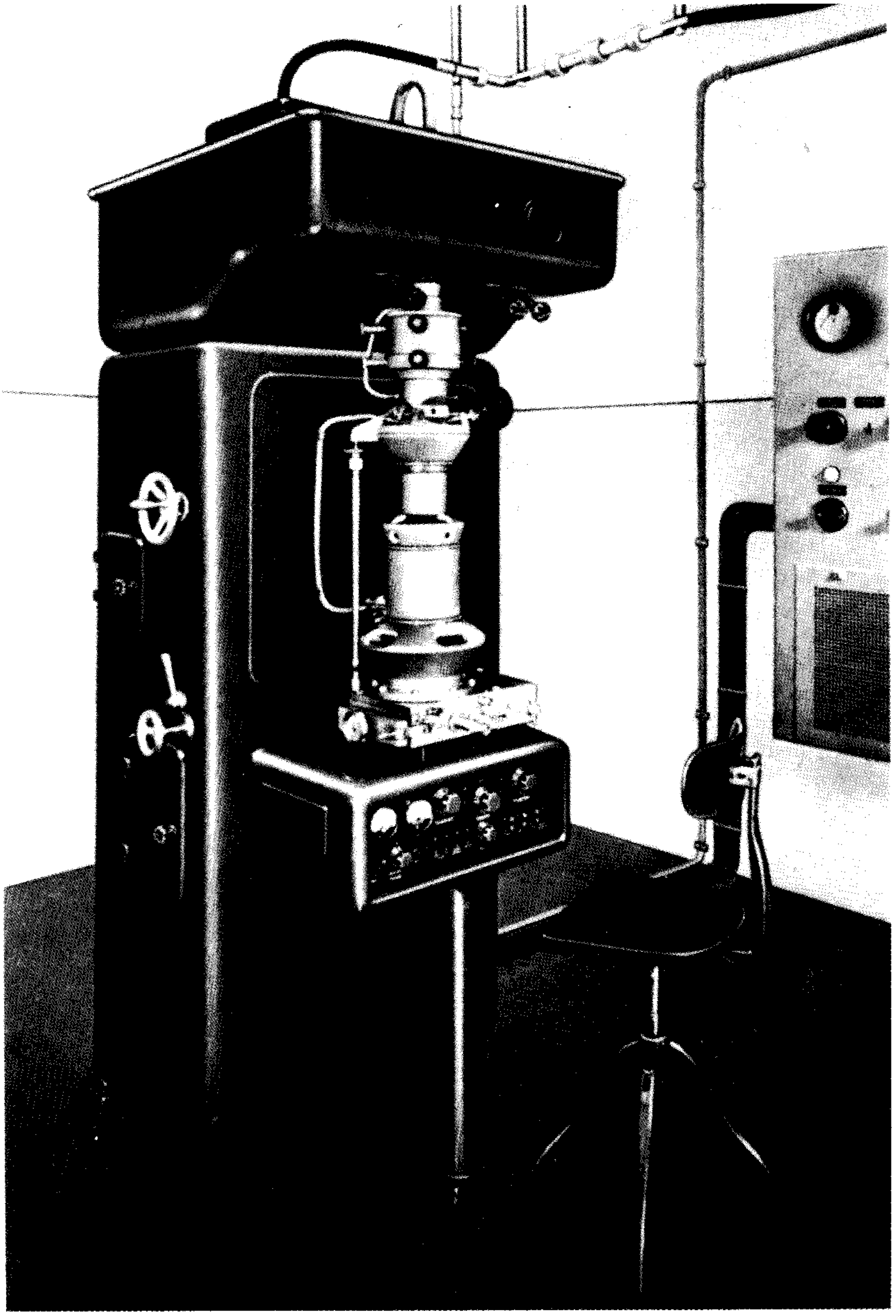


Fig. *11:* The first serially produced electron microscope, by Siemens. General view [Zl]

G. Development of electron microscopy after 1945

Our laboratory had to be reconstructed completely. I could start working with mainly new coworkers as early as June 1945. In spite of difficult conditions in Berlin and Germany, newly developed electron microscopes [22] could be delivered by the end of 1949. In 1954 Siemens had regained its former leading position with the "Elmiskop" [23] (Fig. 12 and Fig. 13). This instrument had, for the first time, two condenser lenses allowing thermal protection of the specimen by irradiating only the small region that is required for the desired final magnification. Since now, for a final magnification of 100 000 x a specimen field of only 1 pm must be irradiated for an image of 10 cm diameter in contrast to earlier irradiation areas of about 1 mm diameter, the power of the electron beam converted into heat in the object can be reduced down to the millionth part. The specimens are heated up just to the extent that the heat power produced can be radiated into the entire region around the object. If the heat power is low, a lower temperature rise with respect to the environment results.

The new instrument was, however, a big disappointment at first when we realized that at this “small region radiation” the image of the specimen fields, which was now no longer hot, became so dark within seconds that all initially visible details disappeared. Investigations then showed that minor residual gases in the evacuated instrument, particularly hydrocarbons, condensed on the cold inner planes of the instrument, i.e. they now even condensed on the specimen itself. The image of the resulting C layer in the irradiated specimen field becomes darker with increasing thickness of the layer. Happily, also this hurdle could, after some time, be surmounted by relatively simple means: The entire environment of the specimen was cooled by liquid air so that the specimen was still markedly warmer than its environment, even without being heated up by the beam. Thus, the residual gases of hydrocarbons condensed on the low-cooled planes and no longer on the specimen.

Along with the successful solution of this problem, another difficulty, that of specimen thickness, had also surprisingly been overcome by newly developed “ultramicrotomes”. Instead of the ground steel knives whose blades were not sufficiently smooth due to crystallization, glass fracture edges were used which had no crystalline unevenness. The usual mechanical translation of the material perpendicular to the knife is-because of mechanical backlash or even oil layers-not sufficiently precise for the desired very small displacements of $\sim 10^{-5}$ mm. Smallest displacements free of flaws were obtained by thermal extension of a rod at whose ends the specimen to be cut was fastened. In order to keep the extremely thin sections smooth, they were dropped into an alcoholic solution immediately after being cut so that they remained entirely flat. Moreover, more suitable fixing agents had been found for the new cutting techniques. The development of these new ultramicrotomes considerably reduced the limitation in the choice of specimens for electron microscopy. For 25 years now, almost all disciplines furthered by light microscopy have also been able to benefit from electron microscopy.

During the last decades, electron microscopy has been advanced in many

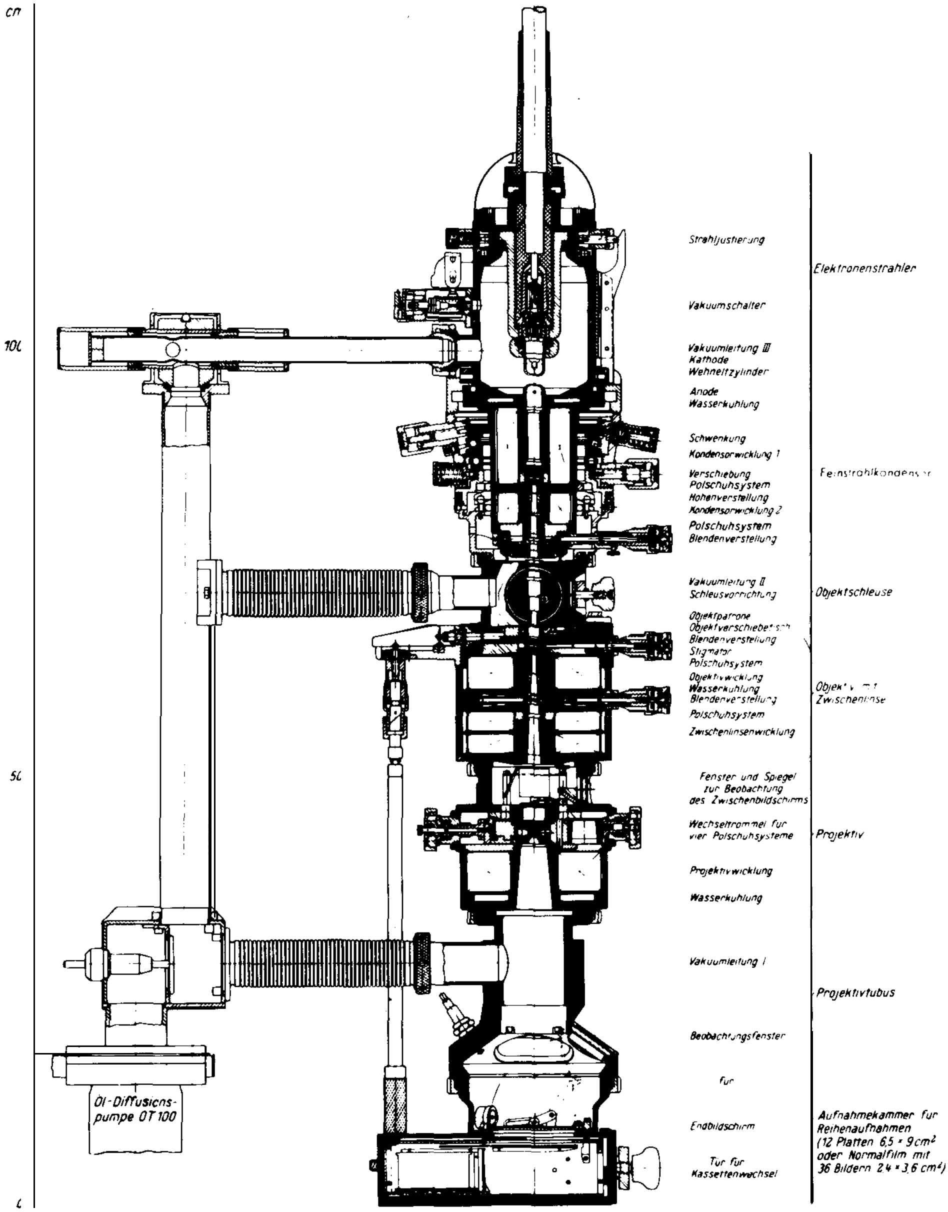


Fig. *12:* The first serially produced 100 kV-Electron microscope with two condenser lenses for "small region radiation" by Siemens. (cross-section) [23].

countries by numerous leading scientists and engineers through new ideas and procedures. I can here only give a few examples: Fig. 14 shows a cross-section through an electron microscope with single-field condenser objective, the specimen being in the field maximum of a magnetic polepiece lens [24]. Thereby, the region of increasing magnetic field in front of the specimen behaves like a condenser of short focal length and the decreasing field region behind the

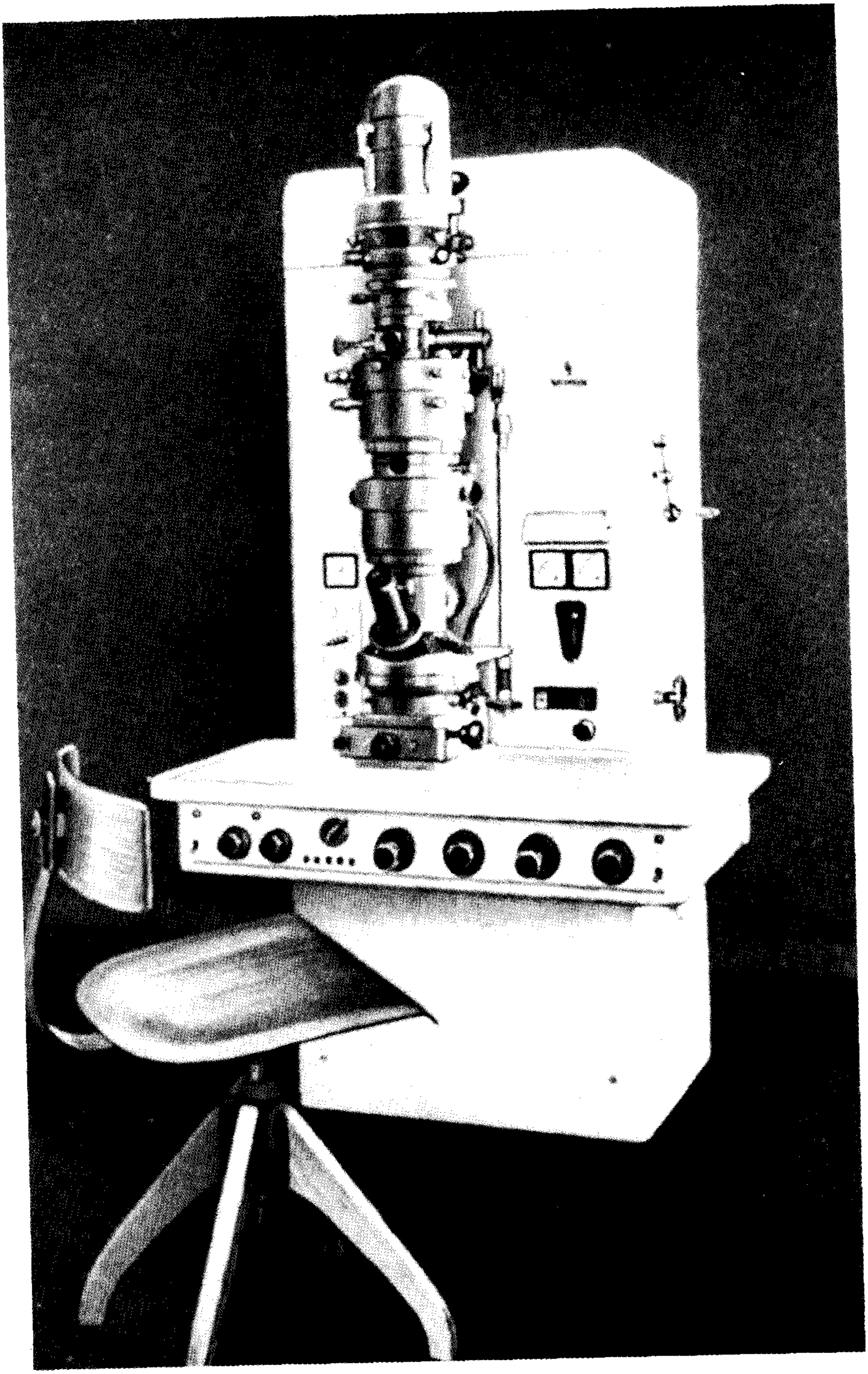


Fig. 13: Same instrument as in Fig. 12 (general view) [23]

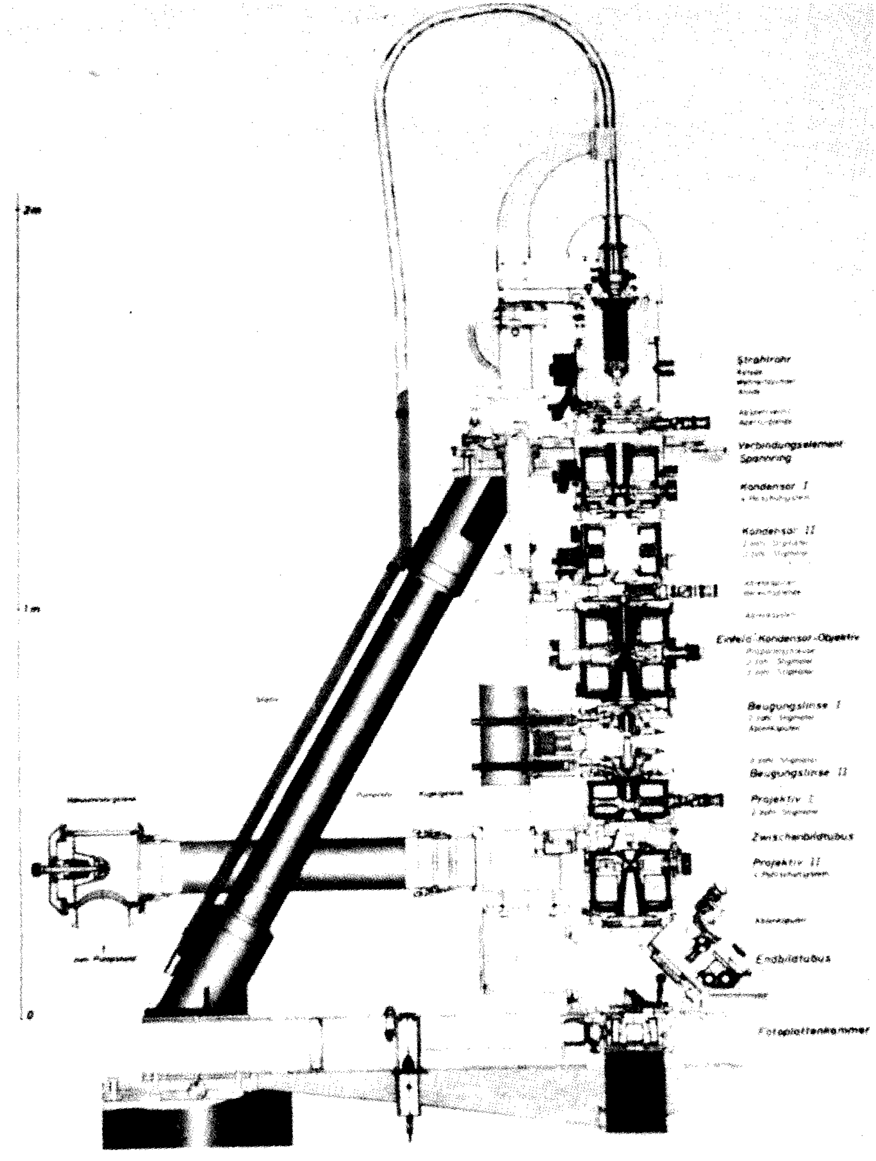


Fig. *14:* Electron microscope with single-field condenser objective.

specimen as an objective of equal focal length. With this arrangement both lenses have a particularly small spherical aberration. Fig. 15 gives a view of the same instrument. Fig. 16 shows an image obtained with this instrument of a platelet of a gold crystal. One can clearly see lattice planes separated by a distance of 1.4 Å. Two such instruments have been further developed in the Institute for Electron Microscopy, which had been set up for me in 1957 by the Max-Planck-Gesellschaft after I had left Siemens. Fig. 17 shows a 3 MV high-voltage instrument developed by Japan Electron Optics Laboratory Co. Ltd. With such instruments whose development was mainly promoted by Gaston

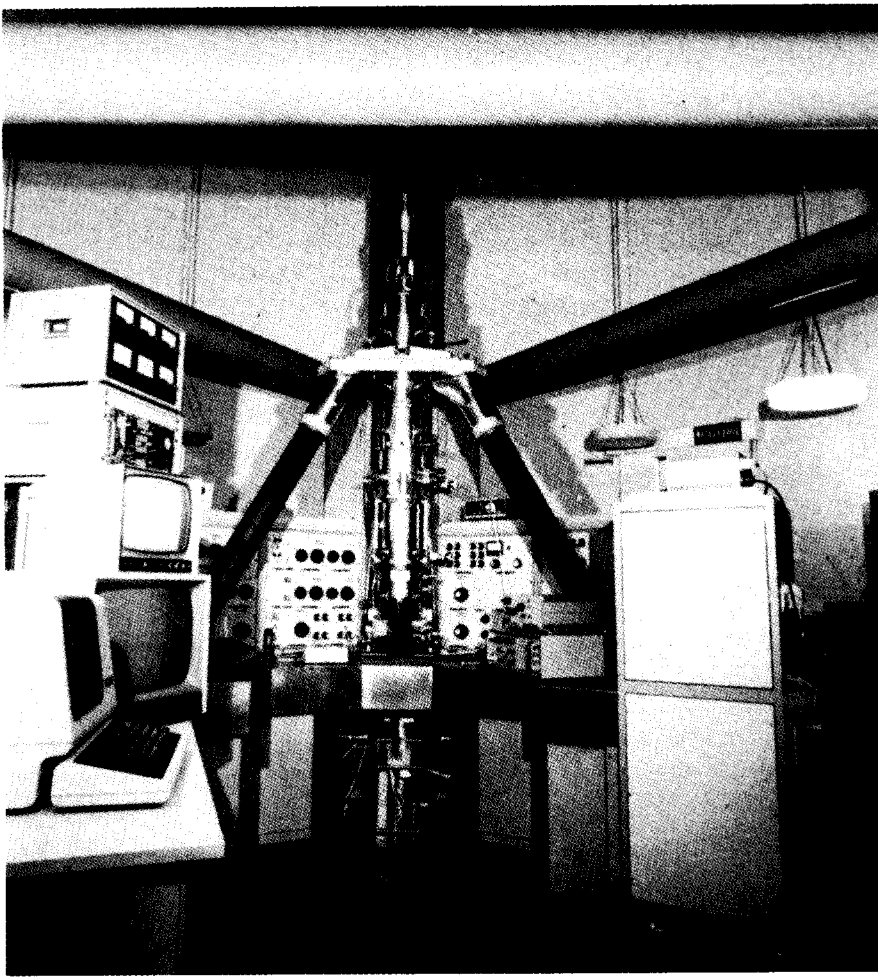


Fig. 15: *Same* instrument as in Fig. 14 (general view) [24].

Dupouy (1900- 1985), apart from extremely high costs, special problems occur in the stabilization of the acceleration voltage and with the protection of the operators against X rays. The aim of the development of these instruments was the investigation of thicker specimens, but now that the problem of stabilizing the high voltages has been overcome, also theresolution has been improved by the shorter material wave length of particularly highly accelerated electrons, so that thinner specimens can also be investigated.

For quite some time now, the cryotechnique-put forward mainly by Fernandez-Moran in the USA-has been of increasing importance. With this technique specimens cooled down to very low temperatures can be studied, because they are more resistant to higher electron doses, i.e. the mobility inside the specimen is very much reduced compared to room temperature. Thus, even after unavoidable ionization, the molecules keep their structure for a long time. In the last years it has been possible to image very beam-sensitive crystals in a cryomicroscope with a resolution of 3.5 Å [25, 26] (Fig. 18) [27].

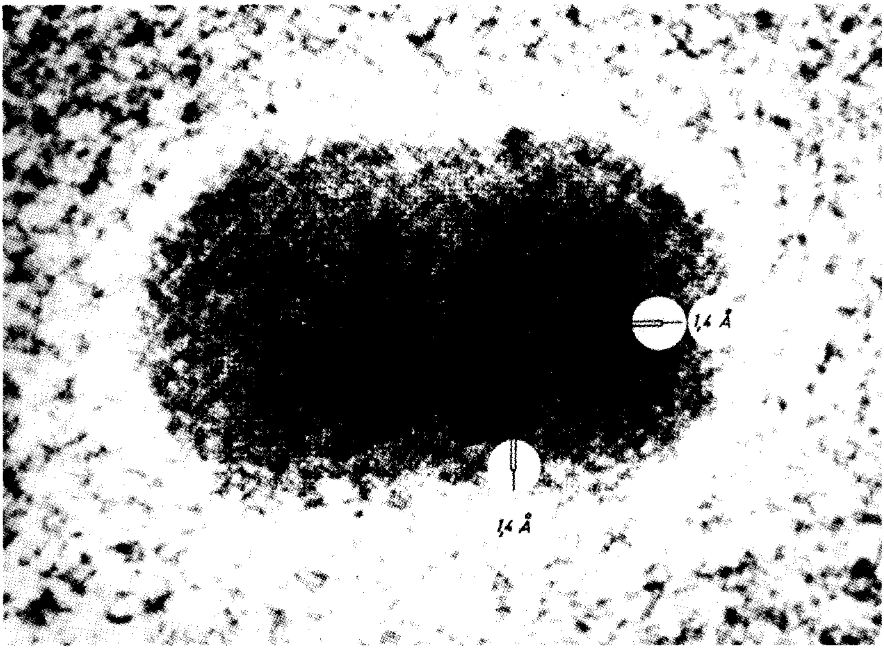


Fig. 16: Plate-like gold crystal, lattice planes with a separation of 0.14 nm, taken with axial illumination.
(U = 100 kV, M_{el} = 800000); taken (1976) by K. Weiss and F. Zemlin with the 100 kV transmission electron microscope with single-field condenser objective at the Fritz-Haber-Institut of the Max-Planck-Gesellschaft.

The specimens were cooled down to -269°C. Direct imaging with sufficient contrast is not possible because the specimen is destroyed at the beam dose needed for normal exposure. Therefore, many very low-dose images are recorded and averaged. Such a single image is very noisy but still contains sufficient periodical information. The evaluation procedure is the following: First, the microgram is digitized using the densitometer so that each image point is given a number which describes the optical density. The underexposed image of the whole crystal is divided like a checkerboard by the computer and then a large number-in our case 400-of these image sub-regions is cross-correlated and summed up by the computer. The resulting image corresponds to a sufficiently exposed micrograph. On the left part in Fig. 18, the initial noisy image of a paraffin crystal is seen; the right side shows the averaged image. Each white point is the image of a paraffin molecule. The long paraffin molecules $C_{44}H_{90}$ are vertical to the image plane. With this procedure electron micrographical images can be processed by the computer. It is even possible to image three-dimensional protein crystals with very high resolution [27]. The computer is a powerful tool in modern electron microscopy.

I cannot go into detail concerning the transmission electron microscopes with electrostatic lenses, the scanning electron microscopes which are widely used mainly for the study of surfaces as well as transparent specimens, the great

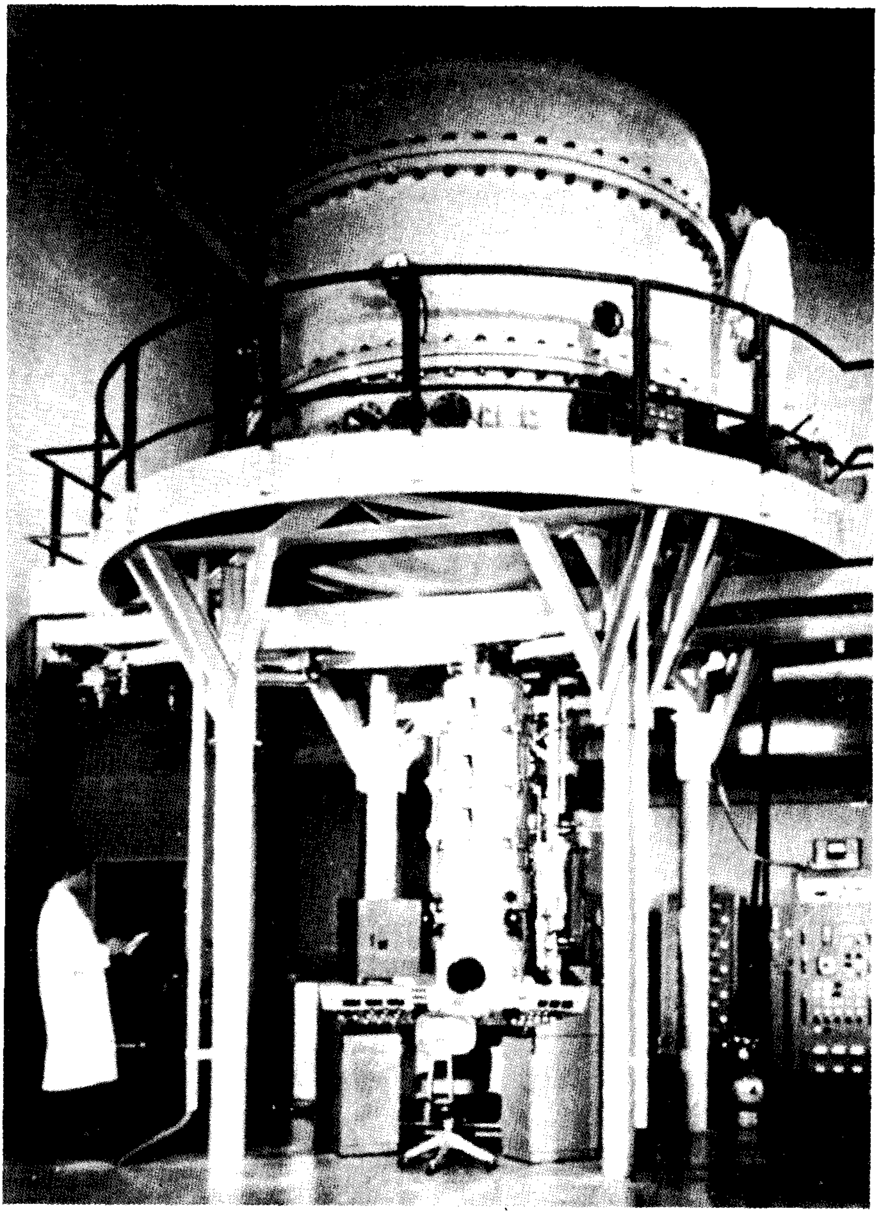


Fig. 17: 1 MV Electron Microscope (Japan Electron Optics Laboratory Co. Ltd.)

importance of various image processing methods carried out partly by the computer, the field-electron microscope and the ion microscope.

The development of the electron microscopy of today was mainly a battle against the undesired consequences of the same properties of electron rays which paved the way for sub-light-microscopical resolution. Thus, for instance, the short material wavelength-prerequisite for good resolution-is coupled with the undesired high electron energy which causes specimen damage. The

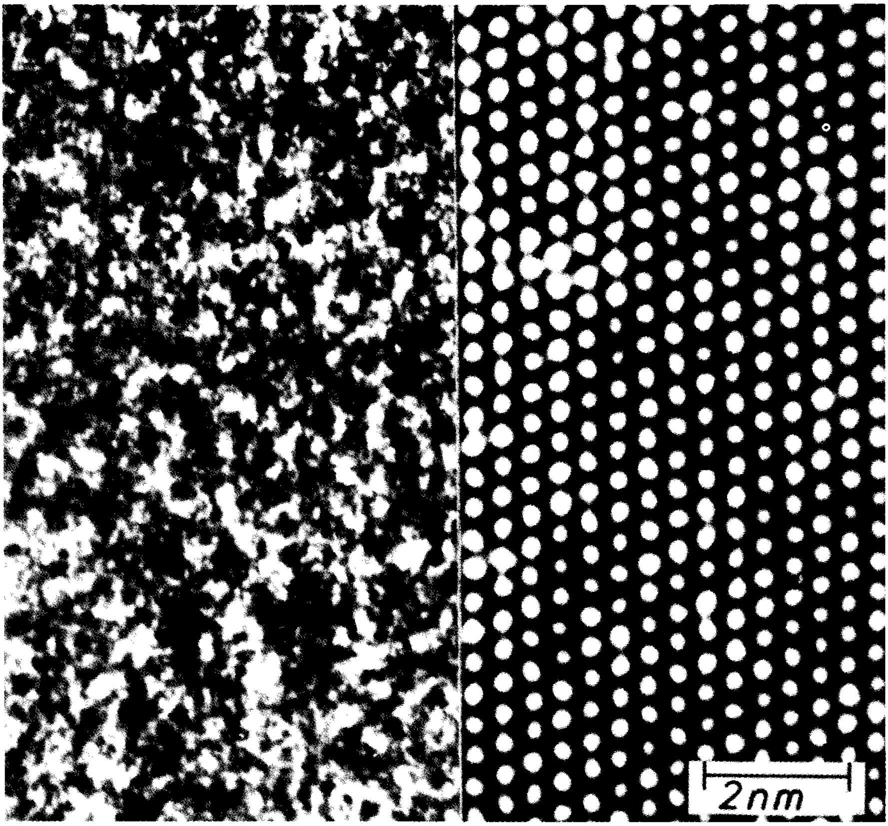


Fig. 18: Paraffin crystal (left: image taken with minimum dose, right: superposition of 400 sub-regions of the left image by means of the computer. [25].

deflectability in the magnetic field, a precondition for lens imaging, can also limit the resolution if the alternating magnetic fields in the environment of the microscope are not sufficiently shielded by the electron microscopy. We should not, therefore, blame those scientists today who did not believe in electron microscopy at its beginning. It is a miracle that by now the difficulties have been solved to an extent that so many scientific disciplines today can reap its benefits.

REFERENCES

1. Rankin, R.: The cathode ray oscillograph. The Electric Club J, II, 620-631 (1905).
2. Hittorf, W.: Uber die Elektrizitatsleitung der Gase, I. und II. (On the electrical conductivity ofgases.) Ann Physik. Chemie, 16, 1-31 und 197-234 (1869), Münster, 9 Oct. 1868.
3. Busch, H.: Über die Wirkungsweise der Konzentrierungsspule bei der Braunschen Röhre. (On the mode of action of the concentrating coil in the Braun tube.) Arch. Elektrotechnik 18, 583-594 (1927), Jena, Physikalisches Institut, März 1927.
4. Ruska, E.: Uber eine Berechnungsmethode des Kathodenstrahloszillographen auf Grund der experimentell gefundenen Abhängigkeit des Schreibfleckdurchmessers von der Stellung der Konzentrierspule. (On a method of designing a cathode ray oscillograph on the basis of the experimentally found dependence of the writing spot diameter on the position of the concentrating coil.) Carried out from 1 November 1928 in the High Voltage Laboratory of the Technische Hochschule Berlin (Director Prof. A. Matthias). Student Project thesis (117 pp.) submitted 10 May 1929.
5. Ruska, E., and Knoll, M.: Die magnetische Sammelspule für schnelle Elektronenstrahlen. (The magnetic concentrating coil for fast electron beams.) Z. techn. Physik 12,389-400 and 448 (1931), submitted 18 April 1931.
6. Knoll, M.: Vorrichtung zur Konzentrierung des Elektronenstrahls eines Kathodenstrahloszillographen. (Devise for concentrating the electron beam of a cathode ray oscillograph.) German Patent NO. 690809, patented on 10 November 1929, granted on 11 April 1940.
7. Ruska, E.: Untersuchung elektrostatischer Sammelvorrichtungen als Ersatz der magnetischen Konzentrierspulen bei Kathodenstrahl-Oszillographen. (Investigation of electrostatic concentrating devices as a substitute for the magnetic concentrating coils in cathode ray oscillographs.) Begun on 18 July 1930 in the High Voltage Laboratory of the Technological University of Berlin (Direktor Prof. Dr. A. Matthias) and submitted on 23 December 1930 as a Diploma Project (pp. l-90).
8. Knoll, M. and Ruska, E.: Beitrag zur geometrischen Elektronenoptik I und II. (Contribution to geometrical electron optics.) Ann. Physik 12, 607-640 and 641-661 (1932), submitted 10 September 1931.
9. Knoll, M.: Berechnungsgrundlagen und neuere Ausführungsformen des Kathodenstrahloszillographen. (The basis of design and new forms of construction of the cathode ray oscillograph.) Manuscript of a lecture in the Cranz-Colloquium at the Technological University of Berlin on 4 June 1931, pp. l-26.
10. Ernst Ruska: “The Early Development of Electron Lenses and Electron Microscopy”, S. Hirzel Verlag Stuttgart (1980), see pp. 113-116.
11. De Broglie, L.: Recherches sur la théorie des quanta. (Researches on the theory of quanta.) These, Paris: Masson & Cie. 1924. Ann. de Physiques 3, 22- 128 (1925).
12. Knoll, M., and Ruska, E.: Das Elektronenmikroskop. (The electron microscope.) Z. Physik 78, 318-339 (1932), submitted 16 June 1932.
13. v. Borries, B., and Ruska, E.: Magnetische Sammellinse kurzer Feldlänge. (Magnetic converging lens of short field length.) German Patent No. 680284, patented on 17 March 1932; Patent granted on 3 August 1939.
14. Ruska, E.: Über ein magnetisches Objektiv fir das Elektronenmikroskop. (On a magnetic objective lens for the electron microscope.) Dissertation of the Technological University of Berlin, submitted 31 August 1933. Z. Physik 89, 90-128 (1934), submitted 5 March 1934.
15. Ruska, E.: Über Fortschritte im Bau und in der Leistung des magnetischen Elektronenmikroskops. (On progress in the construction and performance of the magnetic electron microscope.) Z. Physik 87, 580-602 (1934), submitted 12 Dec. 1933.
16. see 10), pp. 120-122.
17. Marton, L.: La microscope électronique des objects biologiques. (Electron microscopy of biological objects.) Acad. roy. de Belg. Bull de la Cl. des Sci., Ser. 5, 20, 439-446 (1934), Université libre de Bruxells, Mai 1934.

18. Marton, L.: Le microscope électronique. (The electron microscope.) Rev. de Microbiol. appl. 2, 117-124 (1936).
19. see 10), pp. 123-124.
20. Borries, B.v., and Ruska, E.: Vorläufige Mitteilung über Fortschritte im Bau und in der Leistung des Übermikroskops. (Preliminary communication on advances in the construction and performance of the Ultramicroscope.) Wiss. Veröff. a.d. Siemens-Werken 17, 99-106 (1938), submitted 29 Feb. 1938.
21. Borries, B.v., and Ruska, E.: Ein Ubermikroskop für Forschungsindustrie. (An Ultramicroscope for industrial research.) Naturwissenschaften 27, 577-582 (1939), submitted 24 June 1939.
22. Ruska, E.: Über neue magnetische Durchstrahlungs-Elektronenmikroskope im Strahlspannungsbereich von 40...220 kV, Teil I. (0n new magnetic transmission-electron microscopes for beam voltages between 40 and 200 kV.) Kolloid-Zeitschrift 116, 103- 120 (1950), submitted 15 Dec. 1949.
23. Ruska, E., and O. Wolff: Ein hochauflösendes 100-kV-Elektronenmikroskop mit Kleinfelddurchstrahlung (A high-resolution transmission electron microscope (100 kV) with small-area illumination.) Zeitschrift für wissenschaftl. Mikroskopie und mikroskopische Technik 62, 466-509 (1956), submitted 19 July 1955.
24. Riecke, W.D., and Ruska, E.: A 100 kV transmission electron microscope with single-field condenser objective. VI. Int. Congress for Electron Microscopy, Kyoto, Japan, I, 19-20 (1966).
25. Zemlin, F., Reuber, E., Beckmann, E., Zeitler, E. and Dorset, D.L.: Molecular Resolution Electron Micrographs of Monolamellar Paraffin Crystals. Science 229, 461-462 (1985).
26. Dietrich, I., Fox, F., Knapek, E., Lefranc, G., Nachtrieb, K., Weyl, R. and Zerbst, H.: Improvements in electron microscopy by application of superconductivity, Ultramicroscopy 2, 241-249 (197 1).
27. Henderson, R., and Unwin, P.N.T.: Three-dimensional model of purple membrane obtained by electron microscopy, Nature 257, 28-32 (1975).

GERD BINNIG

I was born in Frankfurt, W. Germany, on 7.20., '47 as the first of two sons. My childhood was very much influenced by the Second World War, which had only just ended. We children had great fun playing among the ruins of the demolished buildings, but naturally were too young to realize that much more than just buildings had been destroyed.

Until the age of 31, I lived partly in Frankfurt and partly in Offenbach, a nearby city. I attended school in both cities, and it was in Frankfurt that I started to study physics. Already as a child about 10 years of age, I had decided to become a physicist without actually knowing what it involved. While studying physics, I started to wonder whether I had really made the right choice. Especially theoretical physics seemed so technical, so relatively unphilosophical and unimaginative. In those years, I concentrated more on playing music with friends in a beat-band rather than on physics. My mother had introduced me to classical music very early in life, and I believe this played an important role in my subsequent development. Unfortunately, I started playing the violin rather late, at the age of 15 only, but thoroughly enjoyed being a member of our school orchestra. My brother was responsible for my transition from classics to beat by his perpetually immersing me with the sounds of the Beatles and the Rolling Stones, until I finally really liked that kind of music, and even started composing songs and playing in various beat-bands. In this way, I first learned how difficult teamwork can be, how much fun it is to be creative, and how unpredictable the reaction of an audience can be.

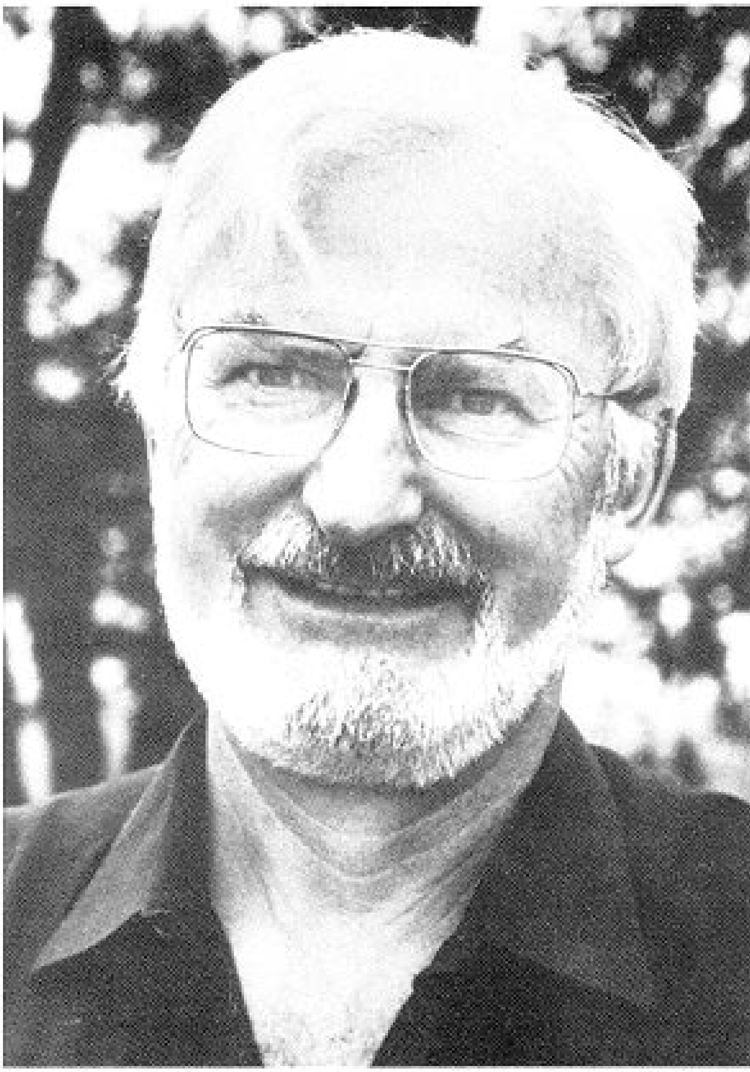
My education in physics gained some significance when I began my diploma work in Prof. Dr. W. Martienssen's group, under Dr. E. Hoenig's guidance. I realized that actually *doing* physics is much more enjoyable than just learning it. Maybe 'doing it' is the right way of learning, at least as far as I am concerned.

I have always been a great admirer of Prof. Martienssen, especially of his ability to grasp and state the essence of the scientific context of a problem. Dr. Hoenig introduced me to experimenting, and exhibited great patience when I asked him very stupid questions in trying to catch up on what I had missed over all the previous years.

In 1969, Lore Wagler became my wife. We had both been studying for quite a long time-Lore is now a psychologist-so only recently did we decide to have children: a daughter born in Switzerland in 1984, and a son born in California in 1986. This was the absolute highlight and most wonderful experience of my whole life. However, fatherhood is not without its sacrifice. For the time being, nearly all my hobbies, like music (singing, playing the guitar and the violin), and sports (soccer, tennis, skiing, sailing and playing golf) have had to take a back seat.

It was in 1978 that Lore-my private psychotherapist-convinced me to accept an offer from the IBM Zurich Research Laboratory to join a physics group. This turned out to be an extremely important decision, as it was here I met Heinrich Rohrer. His way of viewing physics, combined with his humanity and sense of humor, fully restored my somewhat lost curiosity in physics. My years at Rüschlikon, and in IBM Research in general, have been very exciting, not only because of the development of the STM, but also because of the stimulating and pleasant atmosphere created by the people working there, and by those responsible. Working together in a team with Heini Rohrer, Christoph Gerber and Edmund Weibel was an extraordinarily delightful experience, and one for which I shall be eternally grateful. It is also extremely gratifying that our work was recognized far afield. We were first awarded the German Physics Prize, the Otto Klung Prize, the Hewlett Packard Prize, the King Faisal Prize, and now the ultimate crown, the Nobel Prize for Physics. Life certainly does not become easier for a scientist once his work has exceeded a certain significance. But while prizes do add some complications, I must admit they also have their compensations!

(added in *1991)* : In 1990 I joined the Supervisory Board of the Daimler Benz Holding and presently I am involved in a few political activities.



HEINRICH ROHRER

I was born in Buchs, St. Gallen, Switzerland on 6.6., '33 as the third child, half an hour after my twin sister. We were fortunate to enjoy a carefree childhood with a sound mixture of freedom, school and farm work. In 1949, the family moved to Zurich and our way of life changed from country to town. My finding to physics was rather accidental. My natural bent was towards classical languages and natural sciences, and only when I had to register at the ETH (Swiss Federal Institute of Technology) in autumn 1951, did I decide in favor of physics. In the next four years, Professors G. Busch, W. Pauli, and P. Scherrer taught me the rudiments. In autumn 1955, I started work on my Ph.D. Thesis and it was fortuitous that Jörgen Lykke Olsen trusted me to measure the length changes of superconductors at the magnetic-field-induced superconducting transition. He had already pioneered the field with measurements on the discontinuity of Young's modulus. Following in his footsteps, I lost all respect for angstroms. The mechanical transducers were very vibration sensitive, and I learned to work after midnight, when the town was asleep. My four graduate years were a most memorable time, in a group of distinguished graduate students always receptive for fun, and including the interruptions by my basic training courses in the Swiss mountain infantry.

In summer 1961, Rose-Marie Egger became my wife, and her stabilizing influence has kept me on an even keel ever since. Our honeymoon trip led us to the United States where I spent two post-dot years working on thermal conductivity of type-II superconductors and metals in the group of Professor Bernie Serin at Rutgers University in New Jersey. Then in the summer of 1963, Professor Ambros Speiser, Director of the newly founded IBM Research Laboratory in Rüschlikon, Switzerland, made me an offer to join the physics effort there. Encouraged by Bruno Lüthi, who later became a Professor at the University of Frankfurt, and, at the time, strongly recommended the hiring of Gerd Binnig, I accepted to start in December 1963, after having responded to the call of the wild in the form of a four-month camping trip through the USA.

My first couple of years in Rüschlikon were spent studying mainly Kondo systems with magnetoresistance in pulsed magnetic fields. End of the sixties, Keith Blazey interested me to work on $GdAlO_3$, an antiferromagnet on which he had done optic experiments. This started a fruitful cooperation on magnetic phase diagrams, which eventually brought me into the field of critical phenomena. Encouraged by K. Alex Müller, who had pioneered the critical-phenomena effort in our Laboratory, I focused on the bicritical and tetracritical behavior and finally on the random-held problem. These were most enjoyable years, during which so many patient colleagues taught me physics. I left them with some regret, when I ventured with Gerd to discover new shores. We found them. Thank you, Gerd.

In 1974/75, I spent a sabbatical year with Professor Vince Jaccarino and Dr. Alan King at the University of California in Santa Barbara, to get a taste of nuclear magnetic resonance. We solved a specific problem on the bicritical point of MnF_2, their home-base material. We traded experience, NMR and critical phenomena. Rose-Marie and I also took the opportunity at the beginning and end of my sabbatical to show the USA to our two daughters, Doris and Ellen, on two extended camping trips from coast to coast.

In all the years with IBM Research, I have especially appreciated the freedom to pursue the activities I found interesting, and greatly enjoyed the stimulus, collegial cooperation, frankness, and intellectual generosity of two scientific communities, namely, in superconductivity and critical phenomena. I should also like to take this opportunity to thank the many, many friends, teachers, and seniors who have contributed towards my scientific career in any way whatsoever, and most particularly my mother for her unstinting aid and assistance, especially when times were difficult.

SCANNING TUNNELING MICROSCOPY-FROM BIRTH TO ADOLESCENCE

Nobel lecture, December 8, 1986

by

GERD BINNIG AND HEINRICH ROHRER

IBM Research Division, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

We present here the historic development of Scanning Tunneling Microscopy; the physical and technical aspects have already been covered in a few recent reviews and two conference proceedings [l] and many others are expected to follow in the near future. A technical summary is given by the sequence of figures which stands alone. Our narrative is by no means a recommendation of how research should be done, it simply reflects what we thought, how we acted and what we felt. However, it would certainly be gratifying if it encouraged a more relaxed attitude towards doing science.

Perhaps we were fortunate in having common training in superconductivity, a field which radiates beauty and elegance. For scanning tunneling microscopy, we brought along some experience in tunneling [2] and angstroms [3], but none in microscopy or surface science. This probably gave us the courage and light-heartedness to start something which should "not have worked in principle" as we were so often told.

"After having worked a couple of years in the area of phase transitions and critical phenomena, and many, many years with magnetic fields, I was ready for a change. Tunneling, in one form or another had intrigued me for quite some time. Years back, I had become interested in an idea of John Slonczewski to read magnetic bubbles with tunneling; on another occasion, I had been involved for a short time with tunneling between very small metallic grains in bistable resistors, and later I watched my colleagues struggle with tolerance problems in the fabrication of Josephson junctions. So the local study of growth and electrical properties of thin insulating layers appeared to me an interesting problem, and I was given the opportunity to hire a new research staff member, Gerd Binnig, who found it interesting, too, and accepted the offer. Incidentally, Gerd and I would have missed each other, had it not been for K. Alex Müller, then head of Physics, who made the first contacts [l]."

The original idea then was not to build a microscope but rather to perform spectroscopy locally on an area less than 100 Å in diameter.

"On a house-hunting expedition, three months before my actual start at IBM, Heini Rohrer discussed with me in more detail his thoughts on inhomogeneities on surfaces, especially those of thin oxide layers grown on metal

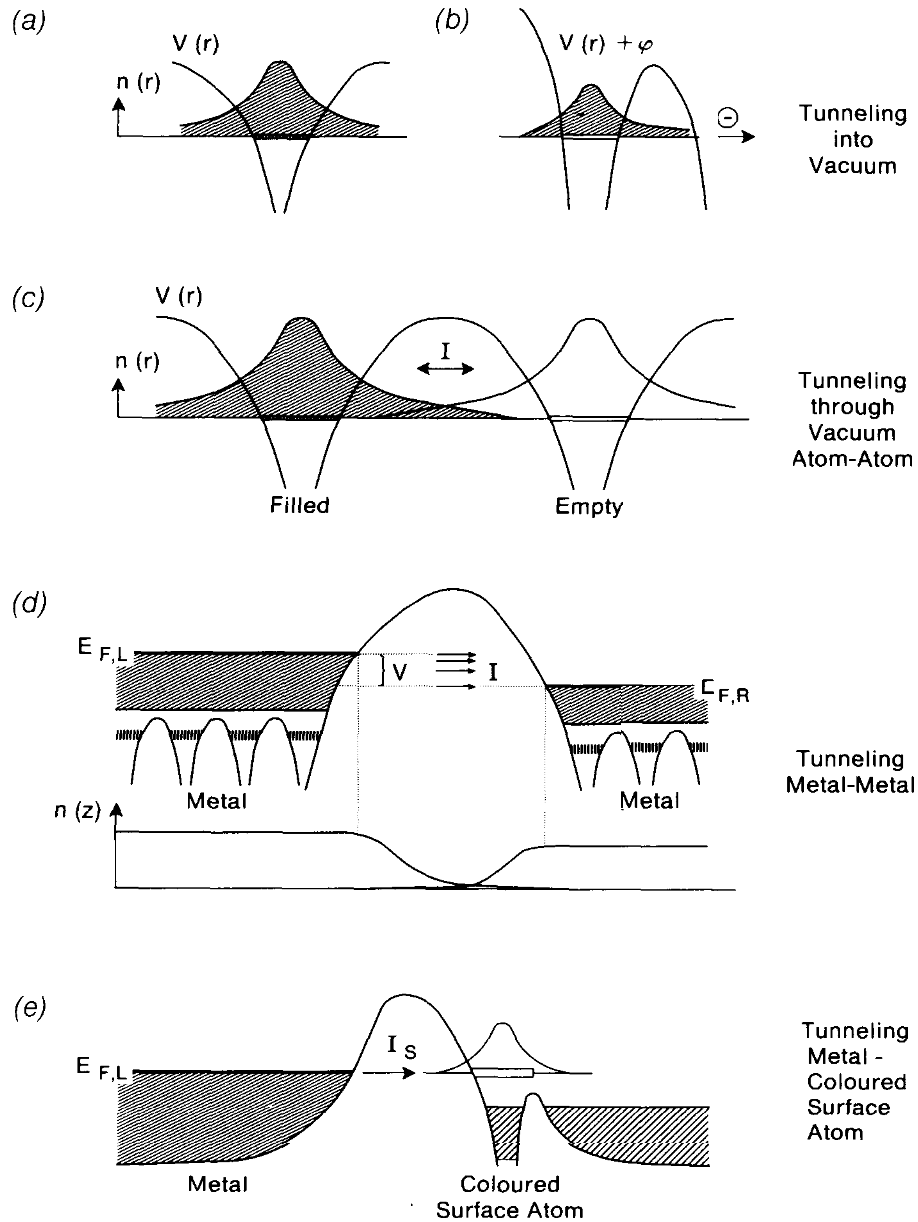


Fig. 1. Tunneling. (a) The wave function of a valence electron in the Coulomb potential well of the atom core plus other valence electrons extends into the vacuum; it "tunnels" into the vacuum. (b) Exposed to an electric field, φ, the electron can tunnel through the potential barrier and leaves the atom. (c) If two atoms come sufficiently close, then an electron can tunnel back and forth through the vacuum or potential barrier between them. (d) In a metal, the potential barriers between the atoms in the interior are quenched and electrons move freely in energy bands, the conduction bands. At the surface, however, the potential rises on the vacuum side forming the tunnel barrier through which an electron can tunnel to the surface atom of another metal close by. The voltage V applied between the two metals produces a difference between the Fermi levels $E_{F,L}$ and $E_{F,R}$, thus providing empty states on the right for the electrons tunneling from the left side. The

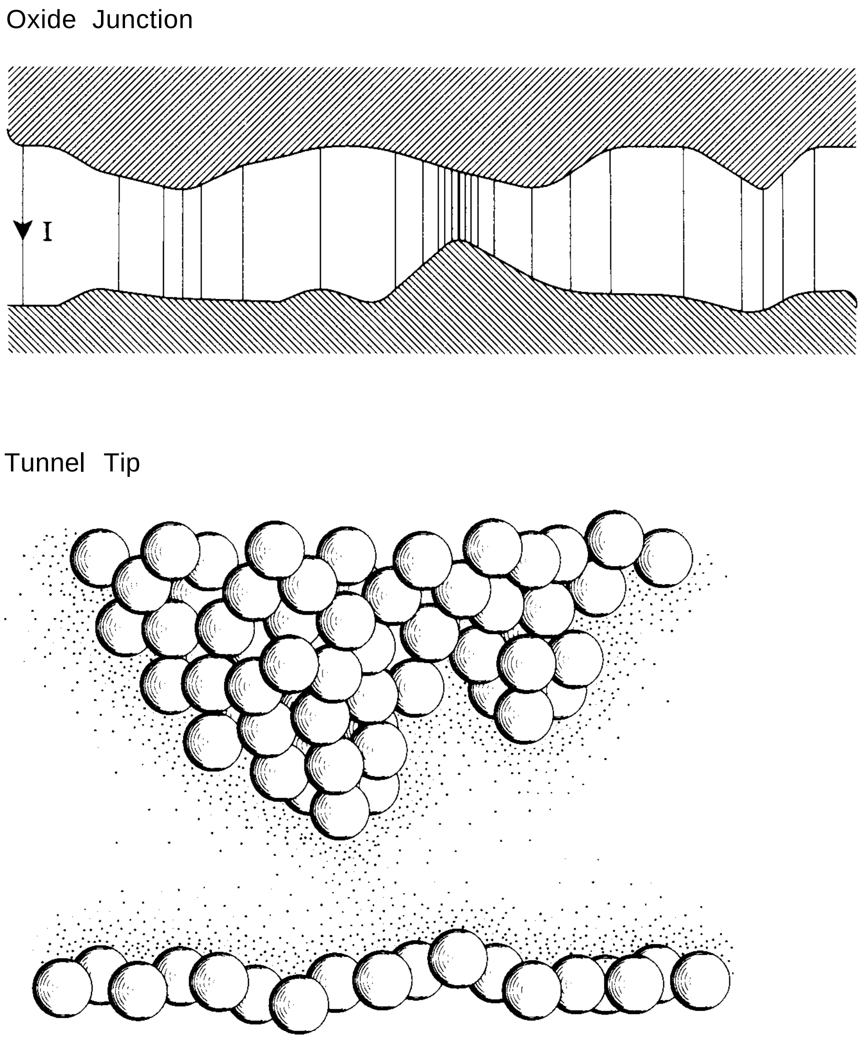


Fig. 2. The principle. The tunneling transmittivity decreases exponentially with the tunneling distance, in vacuum about a factor 10 for every Å. In an oxide tunnel junction, most of the current flows through narrow channels of small electrode separation. With one electrode shaped into a tip, the current flows practically only from the front atoms of the tip, in the best case from a specific orbital of the apex atom. This gives a tunnel-current filament width and thus a lateral resolution of atomic dimensions. The second tip shown is recessed by about two atoms and carries about a million times less current.

---

resulting tunnel current is roughly of the form $I = f(V) \exp(-\sqrt{\varnothing}\, s)$. The f(V) contains a weighted joint local density of states of tip and object, the exponential gives the transmittivity with *f* the averaged tunnel barrier height in eV, and s the separation of the two metals in Å Here f(V) and $\sqrt{\varnothing}$ are material properties obtained by measuring dlnI/dV and dlnI/ds. (e) A simple case of local spectroscopy. A characteristic state, the “color”, of a surface species is observed by the onset of the tunnel-current contribution $I\Sigma$, [see Lang, N. D. (1987) Phys. Rev. Lett. 58, 45, and references therein].

surfaces. Our discussion revolved around the idea of how to study these films locally, but we realized that an appropriate tool was lacking. We were also puzzling over whether arranging tunneling contacts in a specific manner would give more insight on the subject. As a result of that discussion, and quite out of the blue at the LT15 Conference in Grenoble-still some weeks before I actually started at IBM-an old dream of mine stirred at the back of my mind, namely, that of vacuum tunneling. I did not learn until several years later that I had shared this dream with many other scientists, who like myself, were working on tunneling spectroscopy. Strangely enough, none of us had ever talked about it, although the idea was old in principle." Actually, it was 20 years old, dating back to the very beginning of tunneling spectroscopy [4]. Apparently, it had mostly remained an idea and only shortly after we had started, did Seymour Keller, then a member of the IBM Research Division's Technical Review Board and an early advocate of tunneling as a new research area in our Laboratory, draw our attention to W.A. Thompson's attempting vacuum tunneling with a positionable tip [5].

We became very excited about this experimental challenge and the opening up of new possibilities. Astonishingly, it took us a couple of weeks to realize that not only would we have a local spectroscopic probe, but that scanning would deliver spectroscopic and even topographic images, i.e., a new type of microscope. The operating mode mostly resembled that of stylus prolilometry [6], but instead of scanning a tip in mechanical contact over a surface, a small gap of a few angstroms between tip and sample is maintained and controlled by the tunnel current flowing between them. Roughly two years later and shortly before getting our first images, we learned about a paper by R. Young et ***al.*** [7] where they described a type of field-emission microscope they called "topografiner". It had much in common with our basic principle of operating the STM, except that the tip had to be rather far away from the surface, thus on high voltage producing a field-emission current rather than a tunneling current and resulting in a lateral resolution roughly that of an optical microscope. They suggested to improve the resolution by using sharper field-emission tips, even attempted vacuum tunneling, and discussed some of its exciting prospects in spectroscopy. Had they, even if only in their minds, combined vacuum tunneling with scanning, and estimated that resolution they would probably have ended up with the new concept, Scanning Tunneling Microscopy. They came closer than anyone else.

Mid-January 1979, we submitted our first patent disclosure on STM. Eric Courtens, then deputy manager of physics at the IBM Rüschlikon Laboratory, pushed the disclosure to a patent application with "thousands of future STM's". He was the first believer in our cause. Shortly afterwards, following an in-house seminar on our STM ideas, Hans-Jörg Scheel became the third.

For the technical realization of our project, we were fortunate in securing the craftsmanship of Christoph Gerber. "Since his joining IBM in 1966, Christoph had worked with me (HR) on pulsed high-magnetic fields, on phase diagrams, and on critical phenomena. By the end of 1978, we were quite excited about our first experimental results on the random-field problem, but when asked to participate in the new venture, Christoph did not hesitate an instant. He always liked things which were out of the ordinary, and, incidentally, was the second believer. This left me and the random-field problem without his diligent technical support. About a year later, Edi Weibel was the next one to join in, which left another project without technical support. Finally, I completed the team, leaving the random-field problem to others."

During the first few months of our work on the STM, we concentrated on the main instrumental problems and their solutions [8]. How to avoid mechanical vibrations that move tip and sample against each other? Protection against vibrations and acoustical noise by soft suspension of the microscope within a vacuum chamber. How strong are the forces between tip and sample? This seemed to be no problem in most cases. How to move a tip on such a line scale? With piezoelectric material, the link between electronics and mechanics, avoiding friction. The continuous deformation of piezomaterial in the angstrom and subangstrom range was established only later by the tunneling experiments themselves. How to move the sample on a line scale over long distances from the position of surface treatment to within reach of the tip? The 'louse'. How to avoid strong thermally excited length fluctuations of the sample and especially the tip? Avoid whiskers with small spring constants. This led to a more general question, and the most important one: What should be the shape of the tip and how to achieve it? At the very beginning, we viewed the tip as a kind of continuous matter with some radius of curvature. However, we very soon realized that a tip is never smooth because of the finite size of atoms, and because tips are quite rough unless treated in a special way. This roughness implies the existence of minitips as we called them, and the extreme sensitivity of the tunnel current on tip-sample separation then selects the minitip reaching closest to the sample.

Immediately after having obtained the first stable STM images showing remarkably sharp monoatomic steps, we focused our attention onto atomic resolution. Our hopes of achieving this goal were raised by the fact that vacuum tunneling itself provides a new tool for fabricating extremely sharp tips: The very local, high fields obtainable with vacuum tunneling at a few volts only can be used to shape the tip by field migration or by field evaporation. Gently touching the surface is another possibility. All this is not such a controlled procedure as tip sharpening in field-ion microscopy, but it appeared to us to be too complicated to combine STM with field-ion microscopy at this stage. We hardly knew what field-ion microscopy was, to say nothing of working with it. We had no means of controlling exactly the detailed shape of the tip. We repeated our trial-and-error procedures until the structures we observed became sharper and sharper. Sometimes it worked, other times it did not.

But first we had to demonstrate vacuum tunneling. In this endeavor, apart from the occurrence of whiskers, the most severe problem was building vibrations. To protect the STM unit also against acoustical noise, we installed the vibration-isolation system within the vacuum chamber. Our first set-up was designed to work at low temperatures and in ultra-high vacuum (UHV). Low

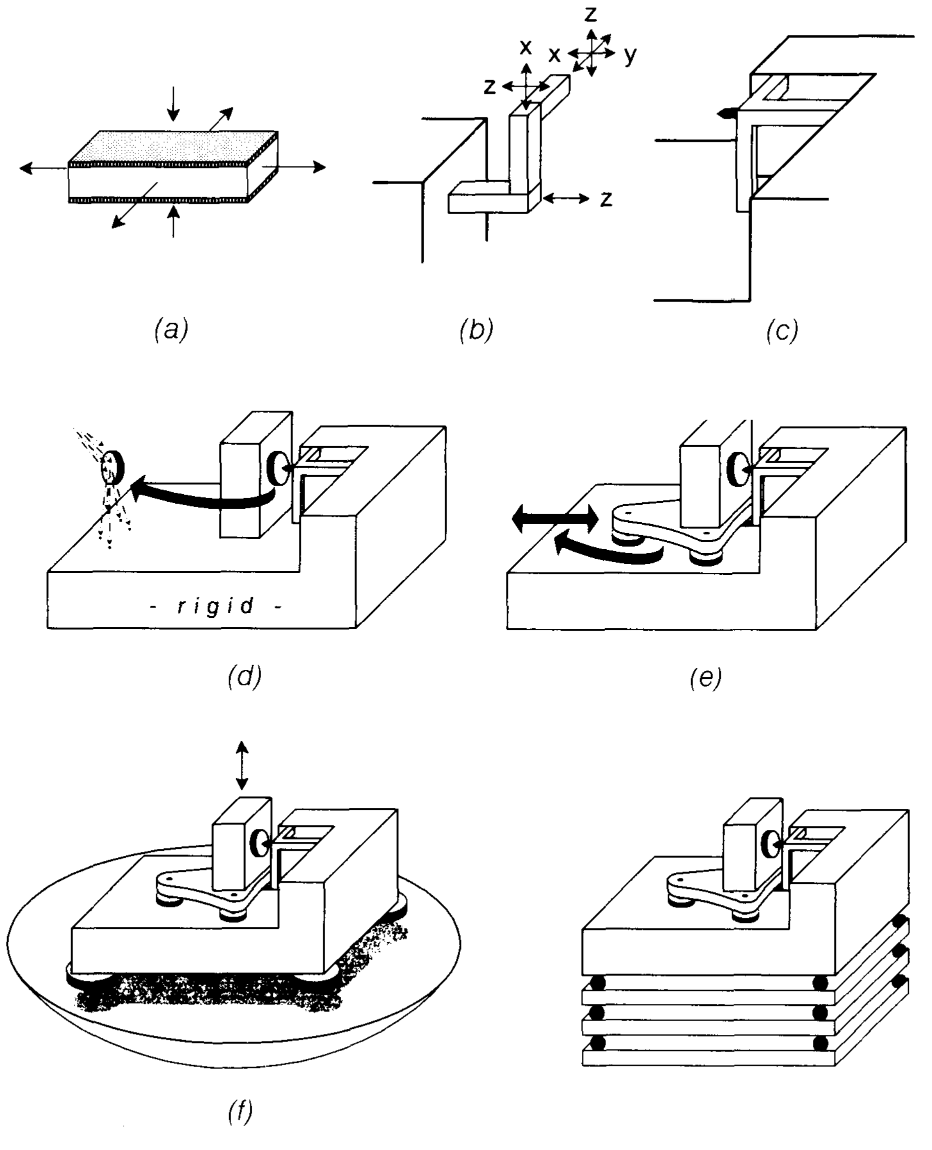


Fig. 3. The instrument. (a) A voltage applied to two electrodes contracts or expands the piezo-electric material in between. The practical total excursion of a piezo is usually in the region of micrometers. (b) A frictionless x-y-z piezodrive, which is quite vibration sensitive. (c) A rigid tripod is at present the piezodrive most used apart from the single-tube scanner. (d) Tripod and sample holder are installed on a rigid frame. The sample has to be cleared from the tip for preparation and sample transfer. (e) Positioning of the sample to within reach of the piezodrive was originally achieved with a piezoelectric 'louse with electrostatically clampable feet. Magnetic-driven positioners and differential screws are also now in use. (f) In the first vibration-isolation system, the tunnel unit with permanent magnets levitated on a superconducting lead bowl. (g) The simple and presently widely used vibration protection with a stack of metal plates separated by viton-a UHV-compatible rubber spacer.

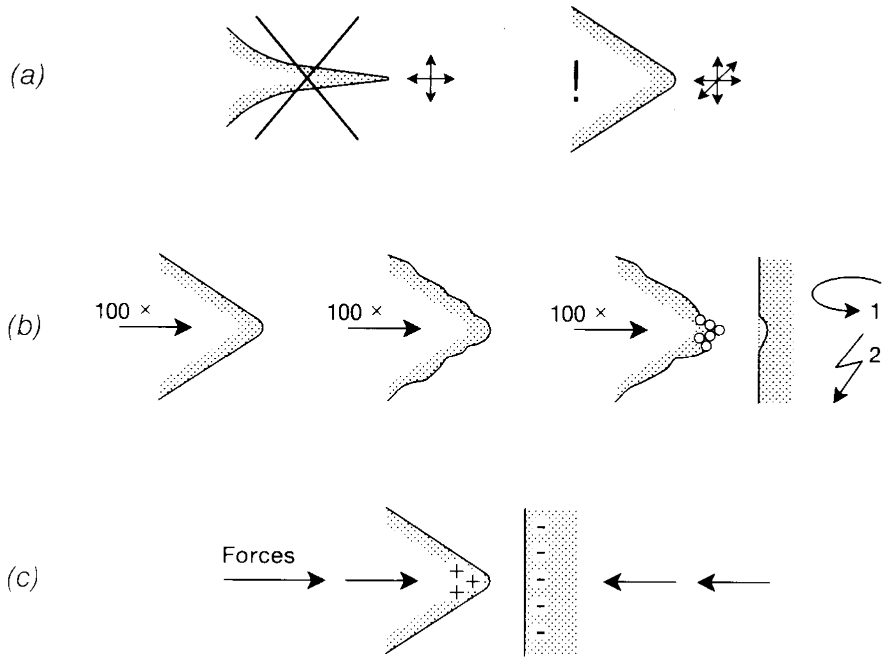


Fig. 4. Tips. (a) Long and narrow tips, or whiskers. arc vibration sensitive and thermally excited. (b) A mechanically ground or etched tip shows sharp minitips. only one of which usually carries the tunnel current. Further sharpening was initially achieved with gentle contact (1), later with field evaporation (2). (c) Electrostatic and interatomic forces between tip and sample do not deform a blunt tip, or a rigid sample, but they make the tunnel gap mechanically unstable when the tip carries a whisker. The response of soft materials like graphite or organic matter to such forces. however. can be appreciable and has to be taken into account

temperatures guaranteed low thermal drifts and low thermal length fluctuations, but we had opted for them mainly because our thoughts were fixed on spectroscopy. And tunneling spectroscopy was a low-temperature domain for both of us with a Ph.D. education in superconductivity. The UHV would allow preparation and retention of well-defined surfaces. The instrument was beautifully designed with sample and tip accessible for surface treatments and superconducting levitation of the tunneling unit for vibration isolation. Construction and first low-temperature and UHV tests took a year. but the instrument was so complicated, wc never used it. We had been too ambitious, and it was only seven years later that the principal problems of a low-temperature and UHV instrument were solved [9]. Instead, we used an exsicator as vacuum chamber, lots of Scotch tape, and a primitive version of superconducting levitation wasting about 20 l of liquid helium per hour. Emil Haupt, our expert glassblower, helped with lots of glassware and, in his enthusiasm, even made the lead bowl for the levitation. Measuring at night and hardly daring to breathe from excitement, but mainly to avoid vibrations, we obtained our first clear-cut exponential dependence of the tunnel current I on tip-sample separation s characteristic for tunneling. It was the portentous night of March 16, 1981.

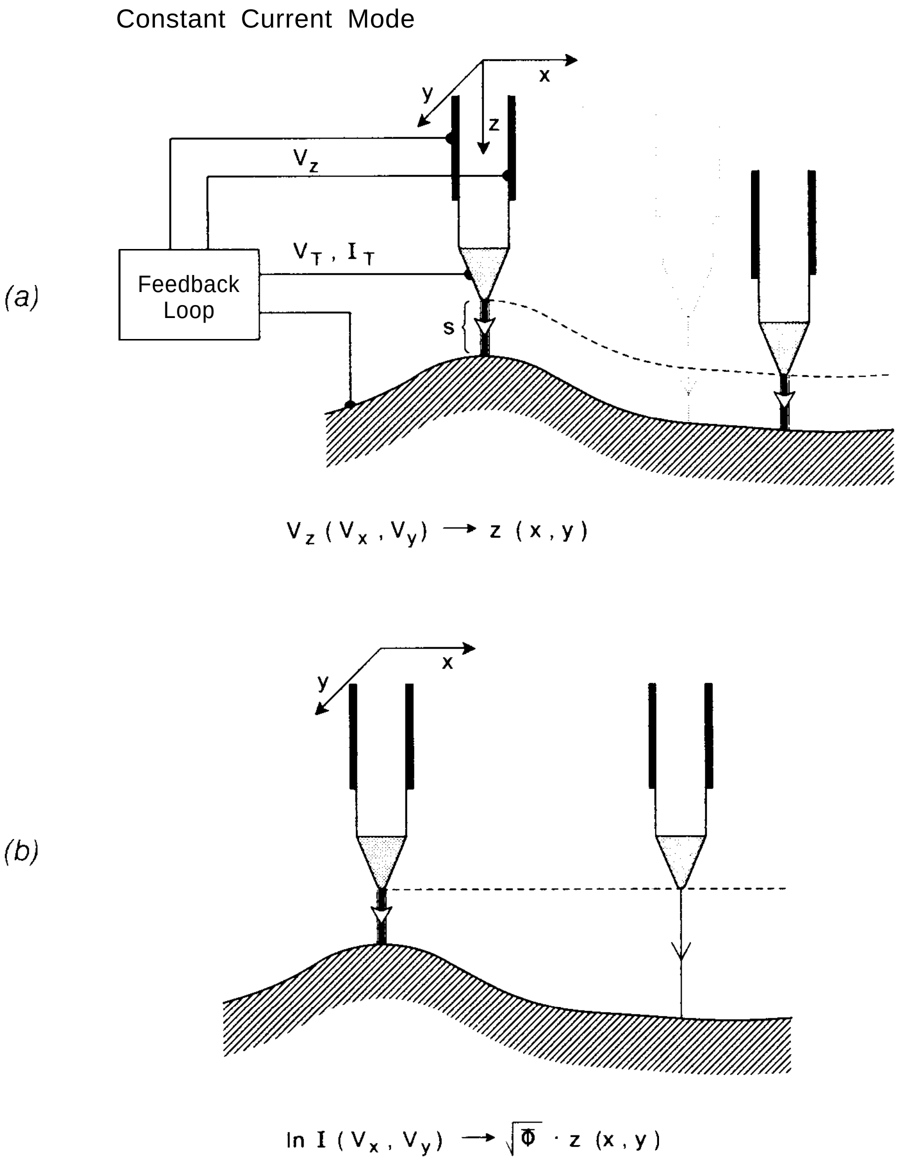


Fig. 5. Imaging. (a) In the constant current mode, the tip is scanned across the surface at constant tunnel current, maintained at a pre-set value by continuously adjusting the vertical tip position with the feedback voltage V_z. In the case of an electronically homogeneous surface, constant current essentially means constant s. (b) On surface portions with denivellations less than a few A-corresponding to the dynamic range of the current measurement-the tip can be rapidly scanned at constant average z-position. Such "current images" allow much faster scanning than in (a) but require a separate determination of $\sqrt{\varnothing}$ to calibrate z. In both cases, the tunnel voltage and/or the z-position can be modulated to obtain in addition, dlnI/dV and/or dlnI/ds, respectively.

So, 27 months after its conception the Scanning Tunneling Microscope was born. During this development period, we created and were granted the necessary elbow-room to dream, to explore, and to make and correct mistakes. We did not require extra manpower or funding, and our side activities produced acceptable and publishable results. The first document on STM was the March/April 1981 in-house Activity Report.

A logarithmic dependence of the tunnel current I on tip-sample separation s alone was not yet proof of vacuum tunneling. The slope of In I versus s should correspond to a tunnel-barrier height of $\phi \approx 5$ eV, characteristic of the average workfunctions of tip and sample. We hardly arrived at 1 eV, indicating tunneling through some insulating material rather than through vacuum. Fortunately, the calibration of the piezosensitivity for small and fast voltage changes gave values only half of those quoted by the manufacturers. This yielded a tunnel-barrier height of more than 4 eV and thus established vacuum tunneling. This reduced piezosensitivity was later confirmed by careful calibration with H.R. Ott from the ETH, Zurich, and of S.Vieira of the Universidad Autónoma, Madrid [10].

U. Poppe had reported vacuum tunneling some months earlier [11], but his interest was tunneling spectroscopy on exotic superconductors. He was quite successful at that but did not measure I(s). Eighteen months later, we were informed that E.C. Teague, in his Thesis, had already observed similar I(s) curves which at that time were not commonly available in the open-literature [12].

Our excitement after that March night was quite considerable. Hirsh Cohen, then Deputy Director of our Laboratory, spontaneously asked us “What do you need?“, a simple and obvious question people only rarely dare to ask. "Gerd immediately wanted to submit a post-deadline contribution [13] to the LT16 Conference to be held in Los Angeles in September. He was going there anyway with his superconducting strontium titanate, and I was sure we would have some topographic STM images by then. And indeed we had. I arranged an extended colloquium tour through the USA for Gerd, but about three weeks before his departure, a friend warned him, that once the news became public, hundreds of scientists would immediately jump onto the STM bandwagon. They did-a couple of years later. After two extended discussions on a weekend hike, he nevertheless became convinced that it was time for the STM to make its public appearance.” Our first attempt to publish a letter failed. “That’s a good sign”, Nico Garcia, a Visiting Professor from the Universidad Autónoma de Madrid, Spain consoled us.

After this first important step with a complete STM set-up, it took us only three months, partly spent waiting for the high-voltage power supplies for the piezcs, to obtain the first images of monosteps [14] on a $CaIrSn_4$ single crystal grown by R. Gambino. Here, the main problem was getting rid of the whiskers we continually created by bumping the tip into the surface. Now we were ready to turn to surface science, first to resolve surface reconstructions. We built a UHV-compatible STM (no longer with Scotch tape!) and as a quick trial, operated it in vacuum suspended from a rubber band. The results indicated that superconducting levitation might be unnecessary.

That was the state of the art for the publicity tour through the USA in September '81. Most reactions were benevolent, some enthusiastic, and two even anticipated the Nobel prize, but the STM was apparently still too exotic for any active outside engagement.

Next, we protected the STM from vibrations by a double-stage spring system with eddy-current damping [8], and incorporated it in a UHV chamber not in use at that moment. We added sputtering and annealing for sample treatment, but no other surface tool to characterize and monitor the state of the sample or tip could yet be combined with that STM. Although the superconducting levitation served for three months only, it was cited for years. It would appear that something complicated is much easier to remember!

A most intriguing and challenging surface-science problem existed, namely, the 7 × 7 reconstruction of the Si(111) surface. A class of fashionable models contained rather rough features which should be resolvable by the STM. So we started to chase after the 7 × 7 structure, and succumbed to its magic. At first, with no success. The STM would function well, sometimes with resolutions clearly around 5 Å, but not our surface preparation. We occasionally found quite nice patterns with monolayer step lines [8] but usually the surface always looked rough and disordered on an atomic scale. One image even foreshadowed the 7 × 7 by a regular pattern of depressions, the precursors of the characteristic corner holes. However, a single event is too risky to make a case for a new structure obtained with a new method. But it boosted our confidence.

By spring '82, STM was already a subject talked about. Supposedly, an image of a vicinal surface expertly prepared with a regular step sequence would have eased the somewhat reserved attitude of the surface-science community. We, however, thought that the mono-, double-, and triple-steps of the $CaIrSn_4$ with atomically flat terraces [14] and the step lines of Si(111) [8] were convincing and promising enough. And instead of wasting further time on uninteresting step lines, we preferred to attack surface reconstructions with known periodicities and with a reasonable chance of learning and contributing something new.

For easier sample preparation and because the demand on resolution was only 8 Å, we changed to a gold single crystal, namely, the (110) surface known to produce a 1 × 2 reconstruction. This seemed to be well within reach of the STM resolution from what we had learned from the silicon step lines. Although some time earlier, we had returned to Karl-Heinz Rieder, the Laboratory's surface-science expert, his Si single crystal in a kind of droplet form, it did not deter him from proposing this gold experiment which meant lending us his Au crystal, and some weeks later we added another droplet to his collection! But in between, with his advice on surface preparation, we succeeded in resolving the 1 × 2 structure [15]. Contrary to expectations, we also had to struggle with resolution, because Au transferred from the surface even if we only touched it gently with our tip. The mobility of Au at room temperature is so high that rough surfaces smooth out after a while, i. e., really sharp Au-coated tips cease to exist. We should like to mention here that later, for measurements on Au(l00), we formed sharp Au tips by field evaporation of Au atoms from sample to tip, and could stabilize them by a relatively high field resulting from a 0.8 V tunnel voltage.

In the case of the Au(110) surface, the atomic resolution was rather a matter of good luck and perseverance. It jumped from high to low in an unpredictable manner, which was probably caused by migrating adatoms on the tip finding a stable position at the apex for a while. We also observed an appreciable disorder leading to long but narrow ribbons of the 1·× 2 reconstruction mixed with ribbons of 1 × 3 and 1× 4 reconstructions and step lines. Nevertheless, these experiments were the first STM images showing atomic rows with atomic resolution perpendicular to the rows. The disorder, intrinsic on this surface, but in its extent criticized from the surface-science point of view, demonstrated very nicely the power of STM as a local method, and about a year later played an important role in testing the first microscopic theories of scanning tunneling microscopy.

With gold, we also performed the first spectroscopy experiment with an STM. We wanted to test a prediction regarding the rectifying I-V characteristic of a sample-tip tunnel junction induced by the geometric asymmetry [16]. Unfortunately, the sample surface became unstable at around 5 V, sample positive, and the small asymmetry observed in this voltage range could also have been due to other reasons. But with reversed polarity, the voltage could be swept up to 20 V producing a whole series of marked resonant surface states [8]. We consider the gold exercise during spring and early summer of '82 a most important step in the development of the method, and the STM had already exceeded our initial expectations. We had also won our first believers outside the Laboratory, Cal Quate from Stanford University [17] and Paul Hansma from the University of California at Santa Barbara [18]. We gave numerous talks on the Au work, and it attracted some attention but all in all, there was little action. We did not even take the time to write a paper- the 7 × 7 was waiting!

Meanwhile, we had also made the first attempts at chemical imaging: Small Au islands on silicon. The islands were visible as smooth, flat hills on a rough surface in the topography, but they were also clearly recognizable as regions with enhanced tunnel-barrier height [8]. Thus, the Au islands were imaged thanks to their different surface electronic properties. It would certainly have been interesting to pursue this line, but we knew that, in principle, it worked,- and the 7 × 7 was still waiting!

We started the second 7 × 7 attempt in autumn 1982 taking into consideration the advice of Franz Himpsel not to sputter the surface. This immediately worked and we observed the 7 × 7 wherever the surface was flat. We were absolutely enchanted by the beauty of the pattern.

"I could not stop looking at the images. It was like entering a new world. This appeared to me as the unsurpassable highlight of my scientific career and therefore in a way its end. Heini realized my mood and whisked me away for

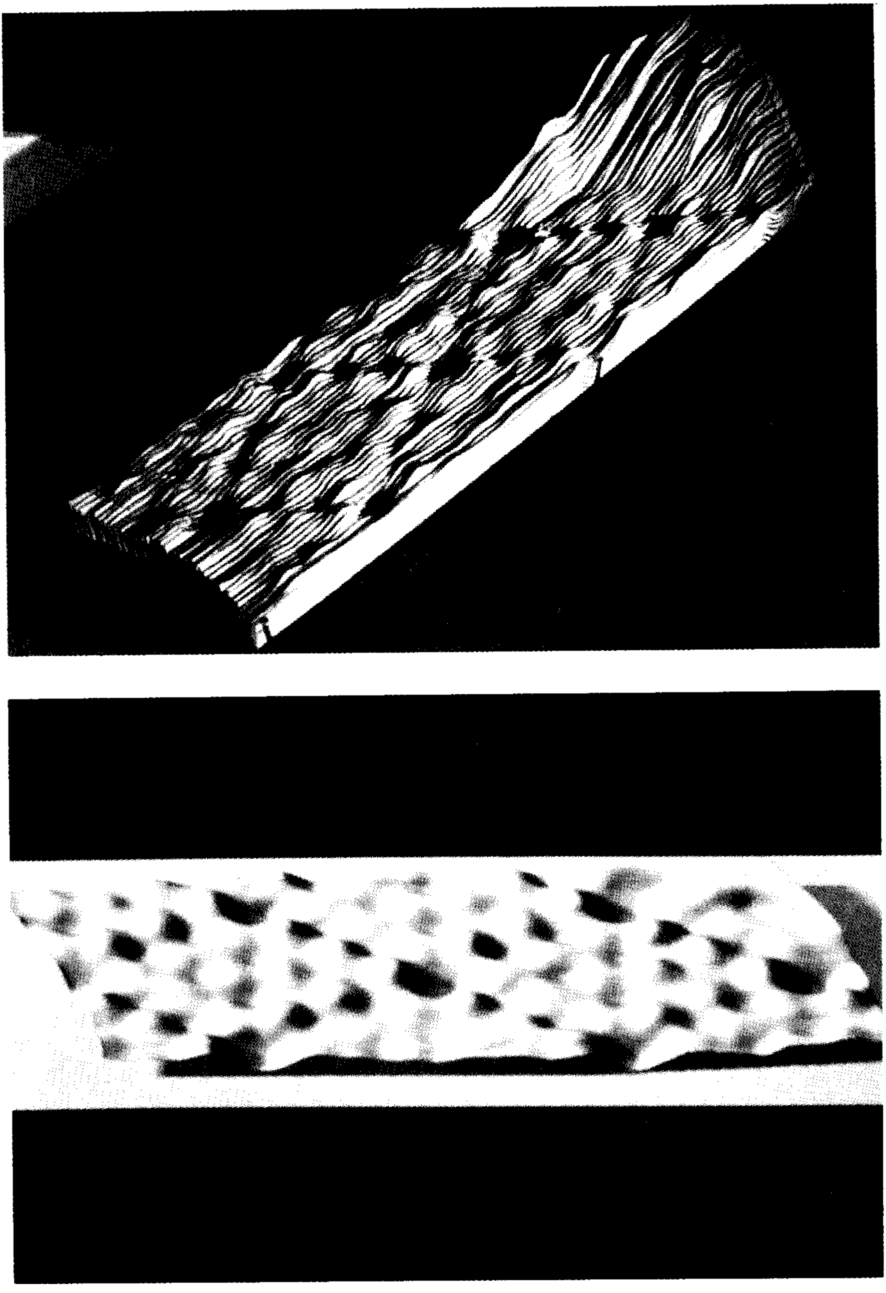


Fig. 6. 7×7 reconstruction of Si(111), (a) Relief assembled from the original recorder traces, from Ref. [19], 0 1983 The American Physical Society, and (h) processed image of the 7×7 reconstruction of Si(111). Characteristic of the rhomhohedral surface unit cell are the corner hole and the 12 maxima, the adatoms. In the processed image, the six adatoms in the right half of the rhombi appear higher. This is an electronic inequivalence on the surface owing to a structural left-right inequivalence in the underlying layers. The reconstruction extends undisturbed to the immediate vicinity of the large "atom hill" on the right.

some days to St. Antönien, a charming village high up in the Swiss mountains, where we wrote the paper on the 7 × 7."

We returned convinced that this would attract the attention of our colleagues, even of those not involved with surface science. We helped by presenting both an unprocessed relief model assembled from the original recorder traces with scissors, Plexiglass and nails, and a processed top view; the former for credibility, the latter for analysis and discussion [19]. It certainly did help, with the result that we practically stopped doing research for a while. We were inundated with requests for talks, and innumerable visitors to our Laboratory were curious to know how to build an STM. However, the number of groups that seriously got started remained small. It seemed there was still a conflict between the very appealing, conceptual easiness of displaying individual atoms in three-dimensional real space direct by recorder traces, and the intuitive reservation that, after all, it just could not be that simple.

Our result excluded all the numerous models that existed, and strangely enough also some that followed. Only one came very close: The adatom model by W. Harrison [20] with just the number of adatoms not quite right. Nowadays, a variation of the adatom model where deeper layers are also reconstructed besides the characteristic 7 × 7 adatom pattern [21], is generally accepted and compatible with most results obtained by various experimental methods like ion channeling [22], transmission electron diffraction [23], and more detailed STM results from other groups [24].

The 7 × 7 experiments also accelerated the first theoretical efforts of STM on a microscopic level. Tersoff and Hamman, and Baratoff [25] applied Bardeen's transfer Hamiltonian formalism to the small geometries of tip and an atomically corrugated surface. Garcia, Ocal, and Flores, and Stoll, Baratoff, Selloni, and Carnevali 'worked out a scattering approach [26]. The two approaches converged; they consoled us by roughly confirming our intuitive view on tunneling in small geometries by simply scaling down planar tunneling, and they certainly improved the acceptance of STM in physics circles. The theoretical treatments concentrated on the nonplanar aspect of tunneling of free electrons, and the STM results on Au(110), still unpublished, served as a testing ground. They remained unpublished for quite some time, since the flashy images of the 7 × 7 silicon surface somehow overshadowed the earlier Au(110) experiments. One reaction to the first attempt to publish them was: "... The paper is virtually devoid of conceptual discussion let alone conceptual novelty . . . I am interested in the behavior of the surface structure of gold and the other metals in the paper. Why should I be excited about the results in this paper? . ." It was certainly bad publication management on our part, but we were not sufficiently familiar with a type of refereeing which searches for weak points, innocently ignoring the essence.

The gold and silicon experiments showed that STM in surface science would benefit greatly from additional, in-situ surface characterization, in particular low-energy electron diffraction (LEED). We had already learned that surfaces, even elaborately prepared, were frequently not as uniform and flat as generally assumed. The in-situ combination of LEED with STM proved extremely helpful, avoiding searching when there was nothing to be searched, and it gave us the opportunity to learn about and work with LEED and Auger electron spectroscopy (AES). The combination of STM with other established surface-science techniques also settled a concern frequently mentioned: How much did our STM images really have in common with surfaces characterized otherwise? We did not share this concern to such a degree, as we had also learned that reconstructions extended unchanged to the immediate vicinity of defect areas, and because we could detect most contaminants or defects individually. Thus, for us, the combined instrumentation was more a practical than a scientific issue.

After a short but interesting excursion with the new STM/LEED/AES combination into resolving and understanding the (100) surface of Au [27], we proceeded into the realms of chemistry. Together with A. Baró, a Visiting Professor from Universidad Autónoma de Madrid, Spain, who also wanted to familiarize himself with the technique, we observed the oxygen-induced 2 × 1 reconstruction of Ni(110) [28], interpreting the pronounced and regularly arranged protrusions we saw as individual oxygen atoms. We had seen atomic-scale features before, which could be interpreted as adsorbates or adsorbate clusters but they were more a nuisance than a matter of interest. The oxygen on Ni experiments demonstrated that the oxygen overlayer was not irreversibly changed by the imaging tunnel tip. This was a most significant result in regard to observing, studying and performing surface chemistry with an STM tip. About a year later, when studying the oxygen-induced 2 × 2 reconstructed Ni(l00) surface, we observed characteristic current spikes which we could attribute to oxygen diffusing along the surface underneath the tip [29]. We noted that the same type of spikes had already been present in our earlier images of oxygen-covered Ni(110), but had been discarded at that time. Not only could diffusing atoms be observed individually, but their migration could be correlated to specific surface features like step lines or bound oxygen atoms, imaged simultaneously. Towards the end of 1983, we also started to probe the possibilities of STM in biology together with H. Gross from the ETH, Zurich. We could follow DNA chains lying on a carbon film deposited on a Ag-coated Si wafer [30].

That year ended with a most pleasant surprise: On Friday December 9, we received a telegram from the secretary of the King Faisal Foundation, followed on Monday by a phone call from the secretary of the European Physical Society announcing the King Faisal Prize of Science and the Hewlett Packard Europhysics Prize, respectively. "The day the telegram arrived, Gerd was in Berlin delivering the Otto Klung Prize lecture. It was also my twentieth anniversary with IBM." This was an encouraging sign that Scanning Tunneling Microscopy was going to make it. It also brought a new flood of requests.

In the summer of 1984, we were finally ready to assume what we had set out to do in autumn 1978, before the notion of microscopy had ever evolved, namely, performing local spectroscopy. Together with H. Fuchs and F. Salvan, we investigated the clean 7 × 7 [1, 3 1] and the $\sqrt{3} \times \sqrt{3}$ Au reconstructions on Si(111) [31], and-right back to the heart of the matter-a thin oxide film on Ni

[1,32]. We could see that surfaces are electronically structured as known, for example, from photoemission experiments, and that we could resolve these electronic structures in space on an atomic scale. We called this (and still do) the color of the atoms. Indeed, the oxide layers were inhomogeneous and most clearly visible in scanning tunneling spectroscopy (STS) images. On the 7 $\times$ 7, we could see by STS down to the second layer, and observe individual dangling bonds between the adatoms [1]. At that time, C. Quate and his group already had an STM running, and they had performed local spectroscopy; not yet with atomic resolution but a low temperature [33]. They had measured the energy gap of a superconductor, and later even plotted its spatial dependence. Spectroscopic imaging was not really surprising, yet it was an important development. We now had the tools to fully characterize a surface in terms of topographic and electronic structure. Although it is usually quite an involved problem to separate the property of interest from a set of STM and STS measurements, our vision of the scanning tunneling microscope had become true. But nevertheless, we heard that this view was not generally shared. Rumors reached us that scientists would bet cases of champagne that our results were mere computer simulations! The bets were probably based on the fact the STM was already three years old, and atomic resolution was still our exclusive property. This was also our concern, but in another way. In late summer ‘83, Herb Budd, promoter of the IBM Europe Institute and an enthusiastic STM supporter, had asked us to run an STM Seminar in summer 1984 within the framework of the Institute. This meant one week with 23 lectures in front of a selected audience of the European academia. At that time, there was no way whatsoever of filling 23 hours, let alone of committing 23 speakers. A year later, we agreed, full of optimism for summer ‘85. In December ‘84, on Cal Quate’s initiative, nine representatives of the most advanced STM groups came together for a miniworkshop in a hotel room in Cancun. It was a most refreshing exchange of ideas, but there was still no other atomic resolution, and thus not a sufficient number of lectures in sight for the Seminar.

In the following few months, the situation changed drastically. R. Feenstra and coworkers came up first with cleaved GaAs [34], C.F. Quate’s group with the 1 $\times$ 1 structure on Pt(l00) [35], and J. Behm, W. Hoesler, and E. Ritter with the hexagonal phase on Pt(l00) [36]. At the American Physical Society March Meeting in 1985, P. Hansma presented STM images of graphite structures of atomic dimensions [37], and when J. Golovchenko unveiled the beautiful results on the various reconstructions of Ge films deposited on Si(111) [38], one could have heard a pin drop in the audience. The atomic resolution was official and scanning tunneling microscopy accepted. The IBM Europe Institute Seminar in July turned into an exclusive workshop for STM’ers, and comprised some 35 original contributions, not all of them on atomic resolution, but already more than in March [39]. “A watershed of ideas” as Cal Quate expressed it.

Our story so far has dealt mainly with the striving for structural and electronic imaging in a surface-science environment with atomic resolution.

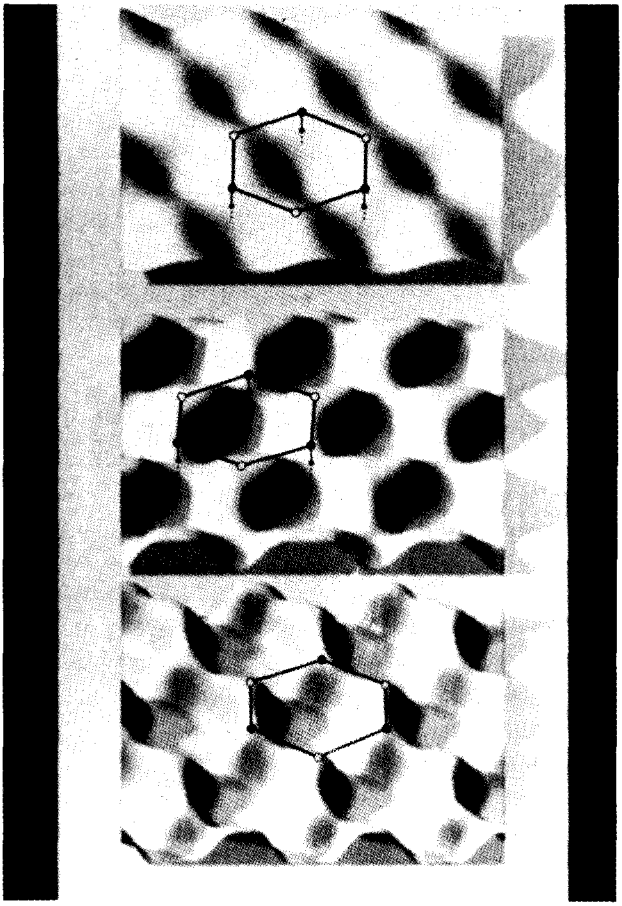


Fig. 7. STM image of cleaved graphite. The top image was taken at a constant tunnel current of 1 nA and at 50 mV. The corrugation traced by the tip reflects the local density of states (LDOS) at the Fermi level and *not* the positions of atoms, which form a flat honeycomb lattice as indicated. The LDOS at the atoms bound to the neighbors in the second layer (open circles) is lower than at the “free” atoms. The image is thus rather a spectroscopic than a topographic one. The middle image is a “current image” showing essentially the same pattern. In the bottom current image, taken closer to the surface, the two inequivalent atoms appear practically identical. This peculiar behavior is compatible with a different local elastic response of the two types of carbon atoms to the interatomic force exerted by the tip compensating for their different LDOS. A local perturbation of the electronic structure might also be important.

Individual atoms had been seen before with field-ion microscopy, and dealt with individually by the atom probe technique [40]. The beauty of these techniques is relativized by the restriction to distinct atom sites on fine tips made from a rather limited selection of materials. Similarly, electron microscopy, the main source of present-day knowledge on submicron structures in practically all areas of science, technology, and industry, has advanced to the atomic level. Imaging of individual atoms or atomic structures, however, is still reserved for specific problems, expertise, and extraordinary equipment. The appeal and the impact of STM lie not only in the observation of surfaces atom by atom, but also in its widespread applicability, its conceptual and instrumental simplicity and its affordability, all of which have resulted in a relaxed and almost casual perception of atoms and atomic structures.

But there are many other aspects, maybe less spectacular but nonetheless significant, which have made STM an accepted and viable method now pursued in many areas of science and technology.

The instruments themselves have become simpler and smaller. Their greatly reduced size allows easy incorporation into other systems, for instance, into a scanning electron microscope [41]. One type of instrument retains accurate sample positioning but is sufficiently rigid for in-situ sample and tip exchange. Other instruments are so rigid they are even insensitive to vibrations when immersed in liquid nitrogen [42], and even small enough to fit through the neck of a liquid-helium storage vessel [43]. These humming-birds of STM, some concepts of which reach back to the squeezable tunnel junctions [18], can also operate at television speed on relatively flat surfaces using single-tube scanners [43, 44]. Also tip preparation has advanced to a level where well-defined pyramidal tips ending with one [45] or more [46] atoms can be fabricated in a UHV environment. Such tips are particularly important for investigations of nonperiodic structures, disordered systems and rough surfaces. They are also interesting in their own right, for example, as low-energy electron and ion point sources.

Outside the physics and surface-science communities, the various imaging environments and imaging capabilities seem as appealing as atomic resolution. Images obtained at ambient-air pressure were first reported in 1984 [47], followed by imaging in cryogenic liquids [42], under distilled water [48], in saline solutions [48], and in electrolytes [49]. Scanning tunneling potentiometry appears to have become an interesting technique to study the potential distribution on an atomic scale of current-carrying microstructures [50]. More recent advances include interatomic-force imaging with the atomic-force microscope [51], with which the structure and elastic properties of conductors and insulators are obtained, and combined imaging of electronic and elastic properties of soft materials [52]. Also the use of spin polarized electron tunneling to resolve magnetic surface structures is being explored.

Finally, we revert to the point where the STM originated: The performance of a local experiment, at a preselected position and on a very small spatial scale down to atomic dimensions. Besides imaging, it opens, quite generally, new possibilities for experimenting, whether to study nondestructively or to modify

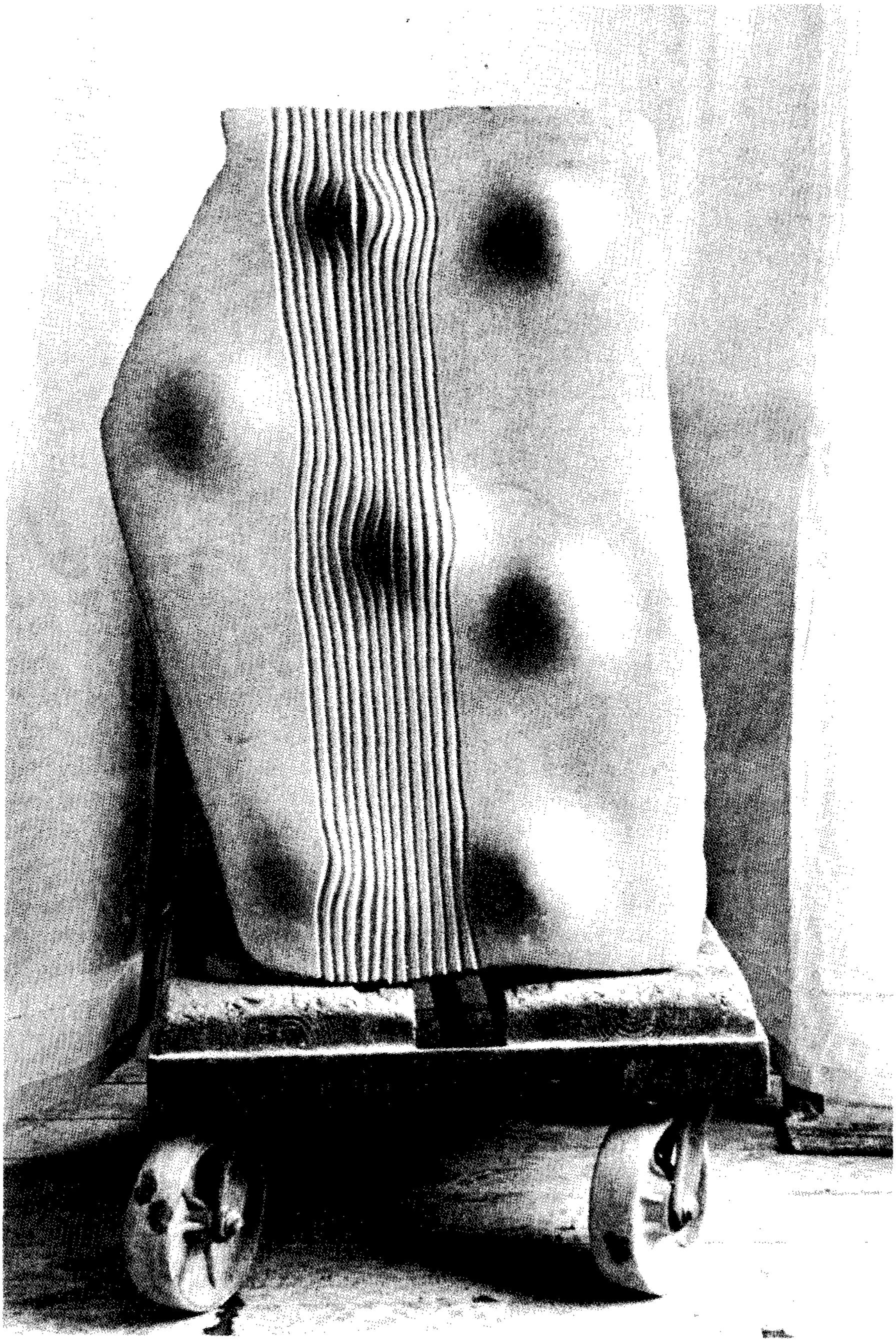


Fig. 8. Artist's conception of spheres. Art and Science are both products of the creativity of Man, and the beauty of nature is reflected in both. Ruedi Rempfler, the sculptor, found his interpretation in the deformation of a surface. It was the tension of the sphere in its environment which fascinated him, more than the mere portrayal of its shape. An independent creation, its visual and conceptual similarity with Fig. 6 is astounding. Original sculpture by Ruedi Rempfler, photograph courtesy of Thomas P. Frey.

locally: Local high electric fields, extreme current densities, local deformations, measurements of small forces down to those between individual atoms, just to name a few, ultimately to handle atoms [53] and to modify individual molecules, in short, to use the STM as a Feynman Machine [54]. This area has not yet reached adolescence.

The STM's "Years of Apprenticeship" have come to an end, the fundamentals have been laid, and the "Years of Travel" begin. We should not like to speculate where it will finally lead, but we sincerely trust that the beauty of atomic structures might be an inducement to apply the technique to those problems where it will be of greatest service solely to the benefit of mankind. Alfred Nobel's hope, our hope, everybody's hope.

ACKNOWLEDGEMENT

We should like to thank all those who have supported us in one way or another, and those who have contributed to the development of Scanning Tunneling Microscopy, and express our appreciation of the pleasant and collegial atmosphere existing in the STM community. Thanks are also due to Dilys Brüllmann for her diligent handling of our manuscripts from the start and for her careful reading of this manuscript, and to Erich Stoll for processing Figs. 6 and 7 using ideas of R. Voss.

REFERENCES

[l] For reviews, see Binnig, G., and Rohrer, H. (1986) IBM J. Res. Develop. 30, 355; Golovchenko, J. A. (1986) Science 232, 48; Behm, R. J., and Hoesler, W. (1986) ***Physics and Chemistry** of **Solid*** Surfaces, Vol VI, (Springer Verlag, Berlin), p. 361; Hansma, P. K., and Tersoff, J. (1987) J. Appl. Phys. ***61,*** RI; Proceedings of the STM Workshops in Oberlech, Austria, July 1-5, 1985, IBM J. Res. Develop. 30, Nos. 4 and 5 (1986); Proceedings of STM '86, Santiago de Compostela, Spain, July 14- 18, 1986, Surface Sci., ***181,*** Nos. 1 and 2 (1987). An article combining technical and biographical details is presented in Dordick, Rowan L. (1986) IBM Research Magazine 24, 2.

[2] Binnig, G. K., and Hoenig, H. E. (1978) Z. Phys. B 32, 23.

[3] Rohrer, H. (1960) Helv. Phys. Acta 33, 675.

[4] Giaever, I. (1974) Rev. Mod. Phys. 46, 245.

[5] Thompson W. A., and Hanrahan, S. F. (1976) Rev. Sci. Instrum. 47, 1303.

[6] Williamson, B. P. (1967) Proc. Inst. Mech. Eng. ***182,21;*** Guenther, K. H., Wierer, P. G., and Bennett, J. M. (1984) Appl. Optics 23, 3820.

[7] Young, R., Ward, J., and Scire, F. (1972) Rev. Sci. Instrum. 43, 999.

[8] For technical details, see Binnig, G., and Rohrer, H. (1982) Helv. Phys. Acta 5.5, 726; ***idem*** (1983) Surface Sci. ***126, 236; idem*** (1985) Sci. Amer. 253, 50.

[9] Marti, 0. (1986) Ph.D. Thesis No. 8095, ETH Zurich, Switzerland; Marti, O., Binnig, G., Rohrer, H., Salemink, H., (1987) Surface Sci., ***181, 230.***

[10] Ott, H. R., and Rohrer, H. (1981) unpublished; Vierira, S. (1986) IBM J. Res. Develop. 30, 553.

[11] Poppe, U. (1981) Verhandl. DPG (VI) ***16, 476.***

[12] Teague, E. C. (1978) Dissertation, North Texas State Univ., Univ. Microfilms International, Ann Arbor, Mich., p. 141; ***idem*** (1978) Bull. Amer. Phys. Soc. 23, 290; ***idem*** (1986) J. Res. Natl. Bur. Stand. ***91,*** 171.

[13] Binnig, G., Rohrer, H., Gerber, Ch., and Weibel, E. (1982) Physica B 109 & ***110, 2075.***

[14] Binnig, G., Rohrer, H., Gerber, Ch., and Weibel, E. (1982) Phys. Rev. Lett. 49,57.

[15] Binnig, G., Rohrer, H., Gerber, Ch., and Weibel, E. (1983) Surface Sci. ***131,*** L379.

[16] Miskowsky, N. M., Cutler, P. H., Feuchtwang, T. E. Shepherd, S. J., Lucas, A. A., and Sullivan, T. E. (1980) Appl. Phys. Lett. 37, 189.

[17] Quate, C. F. (1986) Physics Today 39, 26.

[18] Moreland, J., Alexander, S., Cox, M., Sonnenfeld, R., and Hansma, P. K. (1983) Appl. Phys. Lett. 43, 387. Actually, Paul Hansma was indisposed and could not attend the first seminar given on STM in the USA. However, his students attended, and with them Paul built the squeezable tunnel junction.

[19] Binnig, G., Rohrer, H., Gerber, Ch., and Weibel, E. (1983) Phys. Rev. Lett. 50, 120.

[20] Harrison, W. A. (1976) Surface Sci. 55, 1.

[21] Takayanagi, K., Tanishiro, Y., Takahashi, M., and Takahashi, S. (1985) J. Vac. Sci. Tech. A 3, 1502.

[22] Tromp, R. M., and van Loenen, E. J. (1985) Surface Sci. ***155,441.***

[23] Tromp, R. M. (1985) Surface Sci. ***155,*** 432, and references therein.

[24] Becker, R. S., Golovchenko, J. A., McRae, E. G., and Swartzentruber, B. S. (1985) Phys. Rev. Lett. 55, 2028; Hamers, R. J., Tromp, R. M., and Demuth, J. E. (1986) Phys. Rev. Lett. 56, 1972.

[25] Tersoff, J., and Hamann, D. R. (1983) Phys. Rev. Lett. 50, 1998; Baratoff, A. (1984) Physica ***127B,*** 143.

[26] Garcia, N., Ocal, C., and Flores, F. (1983) Phys. Rev. Lett. 50, 2002; Stoll, E., Baratoff, A., Selloni, A., and Carnevali, P. (1984) J. Phys. C ***17, 3073.***

[27] Binnig; G., Rohrer, H., Gerber, Ch., and Stoll, E. (1984) Surface Sci. ***144,*** 321.

[28] Baró, A. M., Binnig, G., Rohrer, H., **Gerber,** Ch., Stoll, E., Baratoff, A., and Salvan, F. (1984) Phys. Rev. Lett. 52, 1304.
[29] Binnig, G., Fuchs, H., and Stoll, E. (1986) Surface Sci., ***169,*** L295.
[30] Binnig, G., and Rohrer, H. (1984) ***Trends in Physics,*** editors, J. Janta and J. Pantoflicek (European Physical Society) p. 38.
[31] Baratoff, A., Binnig, G., Fuchs, H., Salvan, F., and Stoll, E. (1986) Surface Sci., ***168, 734.***
[32] Binnig, G., Frank, K. H., Fuchs, H., Garcia, N., Reihl, B., Rohrer, H., Salvan, F., and Williams, A. R. (1985) Phys. Rev. Lett. 55, 991; Garcia, R., Saenz, J. J., and Garcia, N. (1986) Phys. Rev. B 33, 4439.
[33] de Lozanne, A. L., Elrod, S. A., and Quate, C. F. (1985) Phys. Rev. Lett. 54, 2433.
[34] Feenstra, R. M., and Fein, A. P. (1985) Phys. Rev. B 32, 1394.
[35] Elrod, S. A., Bryant, A., de Lozanne, A. L., Park, S., Smith, D., and **Quate, C.** F. (1986) IBM J. Res. Develop. 30, 387.
[36] Behm, R. J., Hoesler, W., Ritter, E., and Binnig, G. (1986) Phys. Rev. Lett. 56, 228.
[37] Hansma, P. K. (1985) Bull. Amer. Phys. Soc. 30, 251.
[38] Becker, R. S., Golovchenko, J. A., and Swartzentruber, B. S. (1985) Phys. Rev. Lett. 54, 2678.
[39] Proceedings published (1986) IBM J. Res. Develop. 30, 4/5.
[40] For a review, see Ernst, N., and Ehrlich, G. (1986) *Topicsin Current Physics,* Vol. 40, editor U. Gonser (Springer Verlag, Berlin) p. 75.
[41] Gerber, Ch., Binnig, G., Fuchs, H., Marti, O., and Rohrer, H. (1986) Rev. Sci. Instrum. 57, 221.
[42] Coleman, R. V., Drake, B., Hansma, P. K., and Slough, G. (1985) Phys. Rev. Lett. 55, 394.
[43] Smith, D. P. E., and Binnig, G. (1986) Rev. Sci. Instrum. 57, 2630.
[44] Bryant, A., Smith, D. P. E., and Quate, C. F. (1986) Appl. Phys. Lett. 48, 832.
[45] Fink, H.-W., (1986) IBM J. Res. Develop. 30,460.
[46] Kuk, Y., and Silverman, P. J. (1986) Appl. Phys. Lett. 48, 1597.
[47] Baró, A. M., Miranda, R., Alaman, J., Garcia, N., Binnig, G., Rohrer, H., Gerber, Ch,. and Carrascosa, J. L. (1985) Nature 315, 253.
[48] Sonnenfeld, R., and Hansma, P. K. (1986) Science 232, 211.
[49] Sonnenfeld, R., and Schardt, B. C. (1986) Appl. Phys. Lett. 49, 1172.
[50] Muralt, P., and Pohl, D. W'. (1986) Appl. Phys. Lett. 48, 514.
[51] Binnig, G., Quate, C. F., and Gerber, Ch. (1986) Phys. Rev. Lett. 56,930.
[52] Soler, J. M., Baró, A. M., Garcia, N., and Rohrer, H. (1986) Phys. Rev. Lett. 57, 444; Dürig, U., Gimzewski, J. K., and Pohl, D. W. (1986) Phys. Rev. Lett. 57, 2403.
[53] Becker, R. S., Golovchenko, J. A., and Swartzentruber, B. S. (1987) Nature 325, 419.
[54] Feynman, R. P. (1960) Engr. and Sci., 22, February; Hameroff, S., Schneiker, C., Scott, A., Jablonka, P., Hensen, T., Sarid, D., and Bell, S. (1987).

Physics 1987

J GEORG BEDNORZ and K ALEXANDER MÜLLER

for their important breakthrough in the discovery of superconductivity in ceramic materials

THE NOBEL PRIZE FOR PHYSICS

Speech by Professor GÖSTA EKSPONG of the Royal Academy of Sciences.
Translation from the Swedish text

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen.

The Nobel Prize for Physics has been awarded to Dr. Georg Bednorz and Professor Dr. Alex Müller by the Royal Swedish Academy of Sciences "for their important breakthrough in the discovery of superconductivity in ceramic materials". This discovery is quite recent- less than two years old-but it has already stimulated research and development throughout the world to an unprecedented extent. The discovery made by this year's laureates concerns the transport of electricity without any resistance whatsoever and also the expulsion of magnetic flux from superconductors.

Common experience tells us that bodies in motion meet resistance in the form of friction. Sometimes this is useful, occasionally unwanted. One could save energy, that is to say fuel, by switching off the engine of a car when it had attained the desired speed, were it not for the breaking effect of friction. An electric current amounts to a traffic of a large number of electrons in a conductor. The electrons are compelled to elbow and jostle among the atoms which usually do not make room without resistance. As a consequence some energy is converted into heat. Sometimes the heat is desirable as in a hot plate or a toaster, occasionally it is undesirable as when electric power is produced and distributed and when it is used in electromagnets, in computers and in many other devices.

The Dutch scientist Heike Kamerlingh-Onnes was awarded the Nobel Prize for Physics in 1913. Two years earlier he had discovered a new remarkable phenomenon, namely that the electric resistance of solid mercury could completely disappear. Superconductivity, as the phenomenon is called, has been shown to occur in some other metals and alloys

Why hasn't such an energy saving property already been extensively applied? The answer is, that this phenomenon appears only at very low temperatures; in the case of mercury at -269 degrees Celsius, which means 4 degrees above the absolute zero. Superconductivity at somewhat higher temperatures has been found in certain alloys. However, in the 1970's progress seemed to halt at about 23 degrees above the absolute zero. It is not possible to reach this kind of temperatures without effort and expense. The dream of achieving the transport of electricity without energy losses has been realized only in special cases.

Another remarkable phenomenon appears when a material during cooling crosses the temperature boundary for superconductivity. The field of a nearby magnet is expelled from the superconductor with such force that the magnet can become levitated and remain floating in the air. However, the dream of frictionless trains based on levitated magnets has not been realisable on a large scale because of the difficulties with the necessarily low temperatures.

Dr. Bednorz and Professor Müller started some years ago a search for superconductivity in materials other than the usual alloys. Their new approach met with success early last year, when they found a sudden drop towards zero resistance in a ceramic material consisting of lanthanum-barium-copper oxide. Sensationally, the boundary temperature was 50 % higher than ever before, as measured from absolute zero. The expulsion of magnetic flux, which is a sure mark of superconductivity, was shown to occur in a following publication.

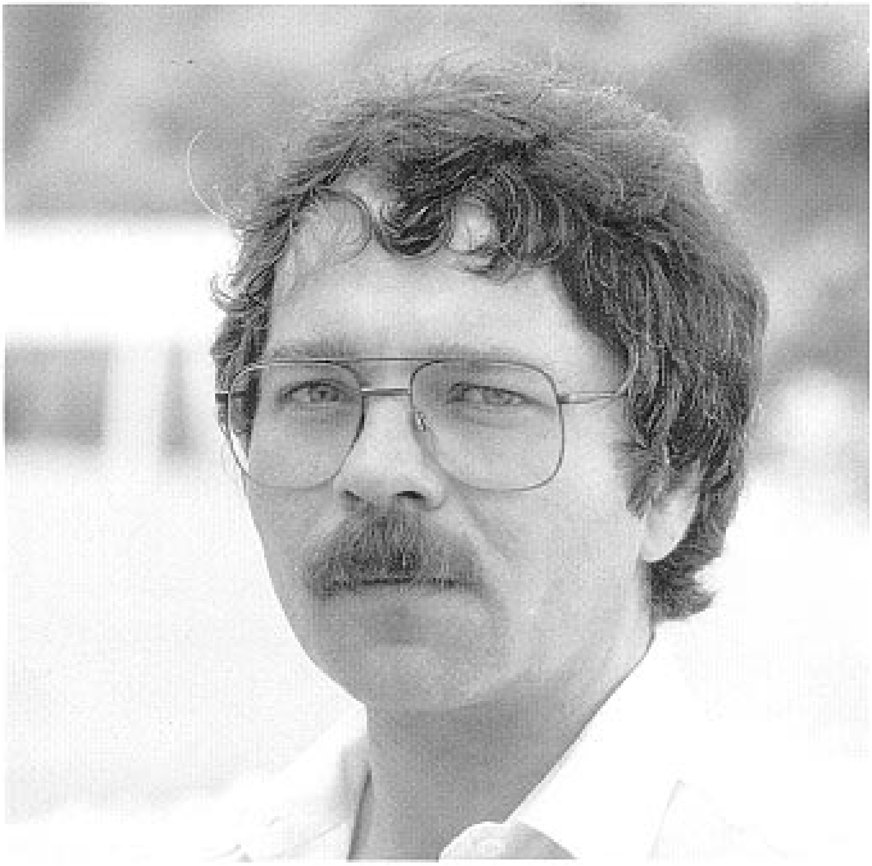
When other experts had overcome their scientifically trained sceptiscism and had carried out their own control experiments, a large number of scientists decided to enter the new line of research. New ceramic materials were synthesized with superconductivity at temperatures such that the cooling suddenly became a simple operation. New results from all over the world flooded the international scientific journals, which found difficulties in coping with the situation. Research councils, industries and politicians are busily considering means to best promote the not so easy development work in order to benefit from the promising possibilities now in sight.

Scientists strive to describe in detail how the absence of resistance to the traffic of electrons is possible and to find the traffic rules, i. e. the laws of nature, which apply. The trio of John Bardeen, Leon Cooper and Robert Schrieffer found the solution 30 years ago in the case of the older types of superconductors and were awarded the Nobel Prize for Physics in 1972. Superconductivity in the new materials has reopened and revitalized the scientific debate in this field.

Herr Dr Bednorz und Herr Professor Müller:

In Ihren bahnbrechenden Arbeiten haben Sic einen neuen, sehr erfolgreichen Weg fir die Erforschung und die Entwicklung der Supraleitung angegeben. Sehr viele Wissenschaftler hohen Ranges sind zurzeit auf dem Gebiet tätig, das Sie eröffnet haben.

Mir ist die Aufgabe zugefallen, Ihnen die herzlichsten Glückwünsche der Küniglich Schwedischen Akademie der Wissenschaften zu übermitteln. Darf ich Sie nun bitten vorzutreten um Ihren Preis aus der Hand Seiner Majestät des Königs entgegenzunehmen.



J. Georg Bednorz

J. GEORG BEDNORZ

I was born in Neuenkirchen, North-Rhine Westphalia, in the Federal Republic of Germany on May 16, 1950, as the fourth child of Anton and Elisabeth Bednorz. My parents, originating from Silesia, had lost sight of each other during the turbulences of World War II, when my sister and two brothers had to leave home and were moved westwards. I was a latecomer completing our family after its joyous reunion in 1949.

During my childhood, my father, a primary school teacher and my mother, a piano teacher, had a hard time to direct my interest to classical music. I was more practical-minded and preferred to assist my brothers in fixing their motorcycles and cars, rather than performing solo piano exercises. At school it was our teacher of arts who cultivated that practical sense and helped to develop creativity and team spirit within the class community, inspiring us to theater and artistic performances even outside school hours. I even discovered my interest in classical music at the age of 13 and started playing the violin and later the trumpet in the school orchestra.

My fascination in the natural sciences was roused while learning about chemistry rather than physics. The latter was taught in a more theoretical way, whereas in chemistry, the opportunity to conduct experiments on our own, sometimes even with unexpected results, was addressing my practical sense.

In 1968, I started my studies in chemistry at the University of Münster, but somehow felt lost due to the impersonal atmosphere created by the large number of students. Thus I soon changed my major to cristallography, that field of mineralogy which is located between chemistry and physics.

In 1972, Prof. Wolfgang Hoffmann and Dr. Horst Böhm, my teachers, arranged for me to join the IBM Zurich Research Laboratory for three months as a summer student. It was a challenge for me to experience how my scientific education could be applied in reality. The decision to go to Switzerland set the course for my future. The physics department of which I became a member was headed by K. Alex Müller, whom I met with deep respect. I was working under the guidance of Hans Jörg Scheel, learning about different methods of crystal growth, materials characterization and solid state chemistry. I soon was impressed by the freedom even I as a student was given to work on my own, learning from mistakes and thus losing the fear of approaching new problems in my own way.

After my second visit in 1973, I came to Rüschlikon for six months in 1974 to do the experimental part of my diploma work on crystal growth and characterization of $SrTiO_3$, again under the guidance of Hans Jörg Scheel. The perovskites were Alex Müller's field of interest and, having followed my work, he encouraged me to continue my research on this class of materials.

In 1977, after an additional year in Münster, I joined the Laboratory of Solid State Physics at the Swiss Federal Institute of Technology (ETH) in Zurich and started my Ph.D. thesis under the supervision of Prof. Heini Gränicher and K. Alex Müller. I gratefully remember the time at the ETH and the family-like atmosphere in the group, where Hanns Arend provided a continuous supply of ideas. It was also the period during which I began to interact more closely with Alex and learned about his intuitive way of thinking and his capability of combining ideas to form a new concept.

In 1978, Mechthild Wennemer followed me to Zurich to start her Ph.D. at the ETH, but more importantly to be my partner in life. I had met her in 1974 during our time together at the University of Münster. Since then she has acted as a stabilizing element in my life and is the best adviser for all decisions I make, sharing the up's and down's in an unselfish way.

I completed my work on the crystal growth of perovskite-type solid solutions and investigating them with respect to structural, dielectric and ferroelectric properties, and joined IBM in 1982. This was the end of a ten-year approach which had begun in 1972.

The intense collaboration with Alex started in 1983 with the search for a high-T, superconducting oxide; in my view, a long and thorny but ultimately successful path. We both realized the importance of our discovery in 1986, but were surprised by the dramatic development and changes in both the field of science and in our personal lives.

(added in 1991) :

Honours

Thirteenth Fritz London Memorial Award (1987) Dannie Heineman Prize (1987), Robert Wichard Pohl Prize (1987) Hewlett-Packard Europhysics Prize (1988) The Marcel Benoist Prize (1986), Nobel Prize for Physics (1987), APS International Prize for Materials Research (1988) Minnie Rosen Award, the Viktor Mortiz Goldschmidt Prize and the Otto Klung Prize

K. Alex Müller

K. ALEX MÜLLER

I was born in Basle, Switzerland, on 20th April 1927. The first years of my life were spent with my parents in Salzburg, Austria, where my father was studying music. Hereafter, my mother and I moved to Dornach near Basle to the home of my grandparents, and from there to Lugano in the italian-speaking part of Switzerland. Here, I attended school and thus became fluent in the Italian language.

My mother died when I was eleven years old, and I attended the Evangelical College in Schiers, situated in a mountain valley in eastern Switzerland. I remained there until I obtained my baccalaureate (Matura) seven years later. This means I arrived in Schiers just before the Second World War started, and left just after it terminated. This was indeed quite a unique situation for us youngsters. Here, in a neutral country, we followed the events of the war worldwide, even in discussion groups in the classes. These college years in Schiers were of significance for my career.

The school was liberal in the spirit of the nineteenth century, and intellectually quite demanding. We were also very active in sports, I especially so in alpine skiing. In my spare time, I became quite involved in building radios and was so fascinated that I really wanted to become an electrical engineer. However, in view of my abilities, my chemistry tutor, Dr. Saurer, eventually convinced me to study physics.

At the age of 19, I did my basic military training in the Swiss army. Upon its completion, I enrolled in the famous Physics and Mathematics Department of the Swiss Federal Institute of Technology (ETH) in Zurich. Our freshman group was more than three times the normal size. We were called the "atom-bomb semester", as just prior to our enrollment nuclear weapons had been used for the first time, and many students had become interested in nuclear physics. The basic course was taught by Paul Scherrer and his vivid demonstrations had a lasting effect on my approach to physics. Other courses were in part not as illuminating, so that, despite good grades, I once seriously considered switching to electrical engineering. However, Dr. W. Känzig, responsible for the advanced physics practicum, convinced me to continue. In the later semesters, Wolfgang Pauli, whose courses and examinations I took, formed and impressed me. He was truly a wise man with a deep understanding of nature and the human being. I did my diploma work under Prof. G. Busch on the Hall effect of gray tin, now known as a semimetal, and, prompted by his fine lectures, also became acquainted with modern solid-state physics.

After obtaining my diploma, following my interest in applications, I worked for one year in the Department of Industrial Research (AFIF) of the ETH on the Eidophor large-scale display system. Then I returned to Prof. Busch's group as an assistant and started my thesis on paramagnetic resonance (EPR).

At one point, Dr. H. Gränicher suggested I look into the, at that time, newly synthesized double-oxide $SrTiO_3$. I found and identified the EPR lines of impurity present in Fe^{3+}.

In spring of 1956, just before starting the latter work, Ingeborg Marie Louise Winkler became my wife. She has always had a substantial influence in giving me confidence in all my undertakings, and over the past 30 years has been my mentor and good companion, always showing interest in my work. Our son Eric, now a dentist, was born in the summer of 1957, six months before I submitted my thesis.

After my graduation in 1958, I accepted the offer of the Battelle Memorial Institute in Geneva to join the staff. I soon became the manager of a magnetic resonance group. Some of the more interesting investigations were conducted on layered compounds, especially on radiation damage in graphite and alkali-metal graphites. The general manager in Geneva, Dr. H. Thiemann, had a strong personality, and his ever-repeated words "one should look for the extraordinary" made a lasting impression on me. Our stay in Geneva was most enjoyable for the family, especially for two reasons: the charm of the city and the birth of our daughter Silvia, now a kindergarten teacher.

While in Geneva, I became a Lecturer (with the title of Professor in 1970) at the University of Zurich on the recommendation of Prof. E. Brun, who was forming a strong NMR group. Owing to this lectureship, Prof. A. P. Speiser, on the suggestion of Dr. B. Lüthi, offered me a position as a research staff member at the IBM Zurich Research Laboratory, Rüschlikon, in 1963. With the exception of an almost two-year assignment, which Dr. J. Armstrong invited me to spend at IBM's Thomas J. Watson Research Center in Yorktown Heights, N.Y., I have been here ever since. For almost 15 years, research on $SrTiO_3$ and related perovskite compounds absorbed my interest: this work, performed with Walter Berlinger, concerned the photochromic properties of various doped transition-metal ions and their chemical binding, ferroelectric and soft-mode properties, and later especially critical and multicritical phenomena of structural phase transitions. In parallel, Dr. Heinrich Rohrer was studying such effects in the antiferromagnetic system of $GdAlO_3$. It was an intense and also, from a personal point of view, happy and satisfying time. While I was on sabbatical leave at the Research Center, he and Dr. Gerd Binnig started the Scanning Tunneling Microscope (STM) project. Just before leaving for the USA, I had been involved in the hiring of Dr. Binnig. Upon my return to Rüschlikon, I closely followed the great progress of the STM project, especially as from 1972 onwards, I was in charge of the physics groups.

The desire to devote more time to my own work prompted me to step down as manager in 1985. This was possible because in 1982 the company had honored me with the status of IBM Fellow. The ensuing work is summarized in Georg Bednorz's part of the Lecture. As he describes there, he joined our Laboratory to pursue his diploma work, on $SrTiO_3$ of course! Ever since making his acquaintance, I have deeply respected his fundamental insight into materials, his human kindness, his working capacity and his tenacity of purpose!

(added in 1991) :

Honorary degrees

Doctor of Science, University of Geneva, Switzerland (1987), Faculty of Physics, the Technical University of Munich, Germany (1987) Università degli Studi di Pavia, Italy (1987), University of Leuven, Belgium (1988) Boston University, USA (1988) Tel Aviv University, Israel (1988), the Technical University of Darmstadt, Germany (1988) University of Nice, France (1989) Universidad Politecnica, Madrid, Spain (1989) University of Bochum, Germany (1990) and Università degli Studi di Roma, Italy (1990)

Honours

Foreign Associate Member, the Academy of Sciences, USA (1989) Special Tsukuba Award (1989) Thirteenth Fritz London Memorial Award (1987), Dannie Heineman Prize (1987) Robert Wichard Pohl Prize (1987), Hewlet-Packard Europhysics Prize (1988) Marcel-Benoist Prize (1986) Nobel Prize in Physics (1987), APS International Prize for New Materials Research (1988) and the Minnie Rosen Award (1988)

PEROVSKITE-TYPE OXIDES - THE NEW APPROACH TO HIGH-T, SUPERCONDUCTIVITY

Nobel lecture, December 8, 1987

by

J. GEORG BEDNORZ and K. ALEX MÜLLER

IBM Research Division, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

PART 1: THE EARLY WORK IN RÜSCHLIKON

In our Lecture, we take the opportunity to describe the guiding ideas and our effort in the search for high-T, superconductivity. They directed the way from the cubic niobium-containing alloys to layered copper-containing oxides of perovskite-type structure. We shall also throw some light onto the circumstances and the environment which made this breakthrough possible. In the second part, properties of the new superconductors are described.

The Background

At IBM's Zurich Research Laboratory, there had been a tradition of more than two decades of research efforts in insulating oxides. The key materials under investigation were perovskites like $SrTiO_3$ and $LaAlO_3$, used as model crystals to study structural and ferroelectric phase transitions. The pioneering ESR experiments by Alex Müller (KAM) [1.1] and W. Berlinger on transition-metal impurities in the perovskite host lattice brought substantial insight into the local symmetry of these crystals, i.e., the rotations of the TiO_6 octahedra, the characteristic building units of the lattice.

One of us (KAM) first became aware of the possibility of high-temperature superconductivity in the 100 K range by the calculations of T. Schneider and E. Stoll on metallic hydrogen [1.2]. Such a hydrogen state was estimated to be in the 2-3 Megabar range. Subsequent discussions with T. Schneider on the possibility of incorporating sufficient hydrogen into a high-dielectric-constant material like $SrTiO_3$ to induce a metallic state led, however, to the conclusion that the density required could not be reached.

While working on my Ph.D. thesis at the Solid State Physics Laboratory of the ETH Zurich, I (JGB) gained my first experience in low-temperature experiments by studying the structural and ferroelectric properties of perovskite solid-solution crystals. It was fascinating to learn about the large variety of properties of these materials and how one could change them by varying their compositions. The key material, pure $SrTiO_3$, could even be turned into a superconductor if it were reduced, i.e., if oxygen were partially removed from its lattice [1.3]. The transition temperature of 0.3 K, however, was too low to create large excitement in the world of superconductivity research. Nevertheless, it was interesting that superconductivity occurred at all, because the carrier densities were so low compared to superconducting NbO, which has carrier densities like a normal metal.

My personal interest in the fascinating phenomenon of superconductivity was triggered in 1978 by a telephone call from Heinrich Rohrer, the manager of a new hire at IBM Rüschlikon, Gerd Binnig. With his background in superconductivity and tunneling, Gerd was interested in studying the superconductive properties of $SrTiO_3$, especially in the case when the carrier density in the system was increased. For me, this was the start of a short but stimulating collaboration, as within a few days I was able to provide the IBM group with Nb-doped single crystals which had an enhanced carrier density compared to the simply reduced material. The increase in T_c, was exciting for us. In the Nb-doped samples $n = 2 \times 10^{20}\,cm^{-3}$, the plasma edge lies below the highest optical phonon, which is therefore unshielded [1.4]. The enhanced electron-phonon coupling led to a T_c of 0.7 K [1.5]. By further increasing the dopant concentration, the T_c even rose to 1.2 K, but this transpired to be the limit, because the plasma edge passes the highest phonon. Gerd then lost his interest in this project, and with deep disappointment I realized that he had started to develop what was called a scanning tunneling microscope (STM). However, for Gerd and Heinrich Rohrer, it turned out to be a good decision, as everyone realized by 1986 at the latest, when they were awarded the Nobel Prize in Physics. For my part, I concentrated on my thesis.

It was in 1978 that Alex (KAM), my second supervisor, took an 18-month sabbatical at IBM’s T. J. Watson Research Center in Yorktown Heights, NY, where he started working in the field of superconductivity. After his return in 1980, he also taught an introductory course at the University of Zurich. His special interest was the field of granular superconductivity, an example being aluminum [1.6], where small metallic grains are surrounded by oxide layers acting as Josephson junctions. In granular systems, the T_c's were higher, up to 2.8 K, as compared to pure Al with $T_c = 1.1$ K.

Involvement with the Problem

It was in fall of 1983, that Alex, heading his IBM Fellow group, approached me and asked whether I would be interested in collaborating in the search for superconductivity in oxides. Without hesitation, I immediately agreed. Alex later told me he had been surprised that he hardly had to use any arguments to convince me; of course, it was the result of the short episode of my activities in connection with the superconducting $SrTiO_3$-he was knocking on a door already open. And indeed, for somebody not directly involved in pushing T_c‘s to the limit and having a background in the physics of oxides, casual observation of the development of the increase of superconducting transition temperatures, shown in Figure 1.1, would naturally lead to the conviction that intermetallic compounds should not be pursued any further. This because since 1973 the highest T_c of 23.3 K [1.7] could not be raised. But nevertheless, the fact that superconductivity had been observed in several complex oxides evoked our special interest.

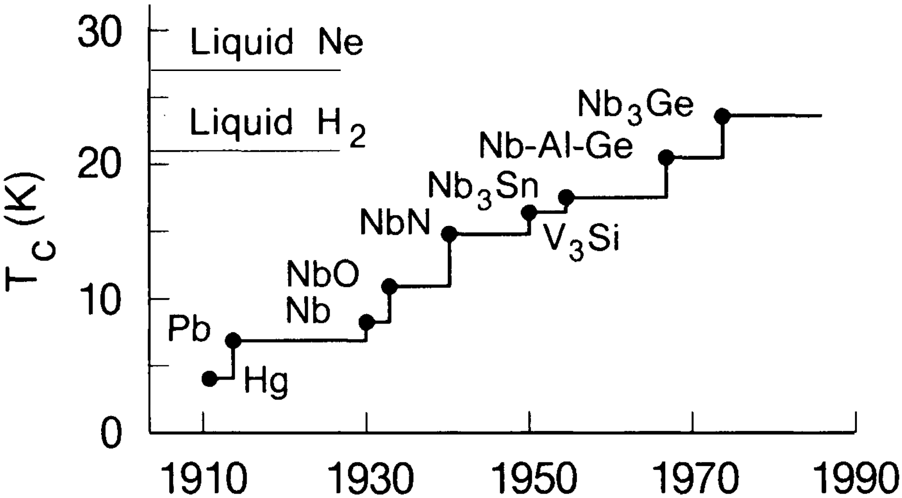


Figure 1.1. Development of the superconducting transition temperatures after the discovery of the phenomenon in 1911. The materials listed are metals or intermetallic compounds and reflect the respective highest T_c's.

The second oxide after $SrTiO_3$ to exhibit surprisingly high T_c's of 13 K was discovered in the Li-Ti-O system by Johnston ***et*** al. [1.8] in 1973. Their multiphase samples contained a $Li_{1+x}Ti_{2-x}O_4$ spinel responsible for the high T,. Owing to the presence of different phases and difficulties in preparation, the general interest remained low, especially as in 1975 Sleight ***et*** al. [1.9] discovered the $BaPb_{1-x}Bi_xO_3$ perovskite also exhibiting a T_c of 13 K. This compound could easily be prepared as a single phase and even thin films for device applications could be grown, a fact that triggered increased activities in the United States and Japan. According to the BCS theory [1.10]

$$k_x T_c = 1.13\hbar\omega_D e^{-1/(N(E_F)) x V^*},$$

both mixed-valent oxides, having a low carrier density $n = 4 \times 10^{21}/cm^3$ and a comparatively low density of states per unit cell $N(E_F)$ at the Fermi level, should have a large electron-phonon coupling constant V*, leading to the high T_c's. Subsequently, attempts were made to raise the T_c in the perovskite by increasing $N(E_F)$ via changing the Pb:Bi ratio, but the compound underwent a metal-insulator transition with a different structure, thus these attempts failed.

We in Rüschlikon felt and accepted the challenge as we expected other metallic oxides to exist where even higher T_c's could be reached by increasing $N(E_F)$ and/or the electron-phonon coupling. Possibly we could enhance the latter by polaron formation as proposed theoretically by Chakraverty [1.11] or

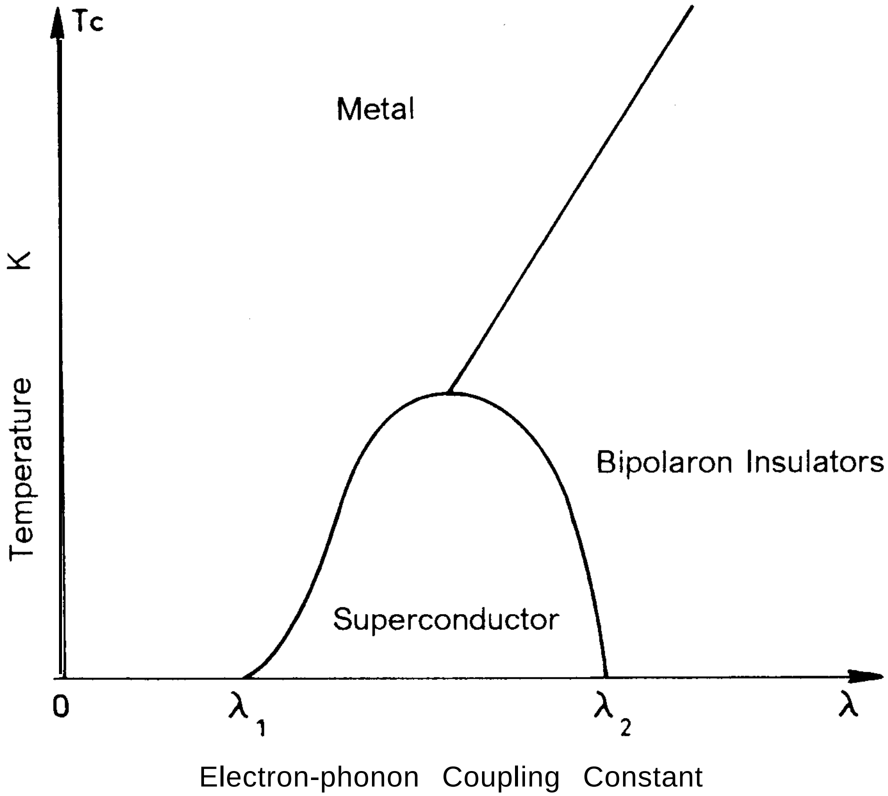


Figure 1.2. Phase diagram as a function of electron-phonon coupling strength. From [1.11], ©Les Editions de Physique 1979.

by the introduction of mixed valencies. The intuitive phase diagram of the coupling constant $\lambda = N(E_F)^x V^*$ versus T proposed by Chakraverty for polaronic contributions is shown in Figure 1.2. There are three phases, a metallic one for small λ and an insulating bipolaronic one for large λ, with a superconductive phase between them, i.e., a metal-insulator transition occurs for large λ. For intermediate λ, a high-T, superconductor might be expected. The question was, in which systems to look for superconductive transitions.

The Concept

The guiding idea in developing the concept was influenced by the Jahn-Teller (JT) polaron model, as studied in a linear chain model for narrow-band intermetallic compounds by Höck *et al.* [1.12].

The Jahn-Teller (JT) theorem is well-known in the chemistry of complex units. A nonlinear molecule or a molecular complex exhibiting an electronic degeneracy will spontaneously distort to remove or reduce this degeneracy. Complexes containing specific transition-metal (TM) central ions with special valency show this effect. In the linear chain model [1.12], for small JT distortions with a stabilization energy E_{JT} smaller than the bandwidth of the metal, only a slight perturbation of the traveling electrons is present. With increasing

E_{JT}, the tendency to localization is enhanced, and for E_{JT} being of the magnitude of the bandwidth, the formation of JT polarons was proposed.

These composites of an electron and a surrounding lattice distortion with a high effective mass can travel through the lattice as a whole, and a strong electron-phonon coupling exists. In our opinion, this model could realize the Chakraverty phase diagram. Based on the experience from studies of isolated JT ions in the perovskite insulators, our assumption was that the model would

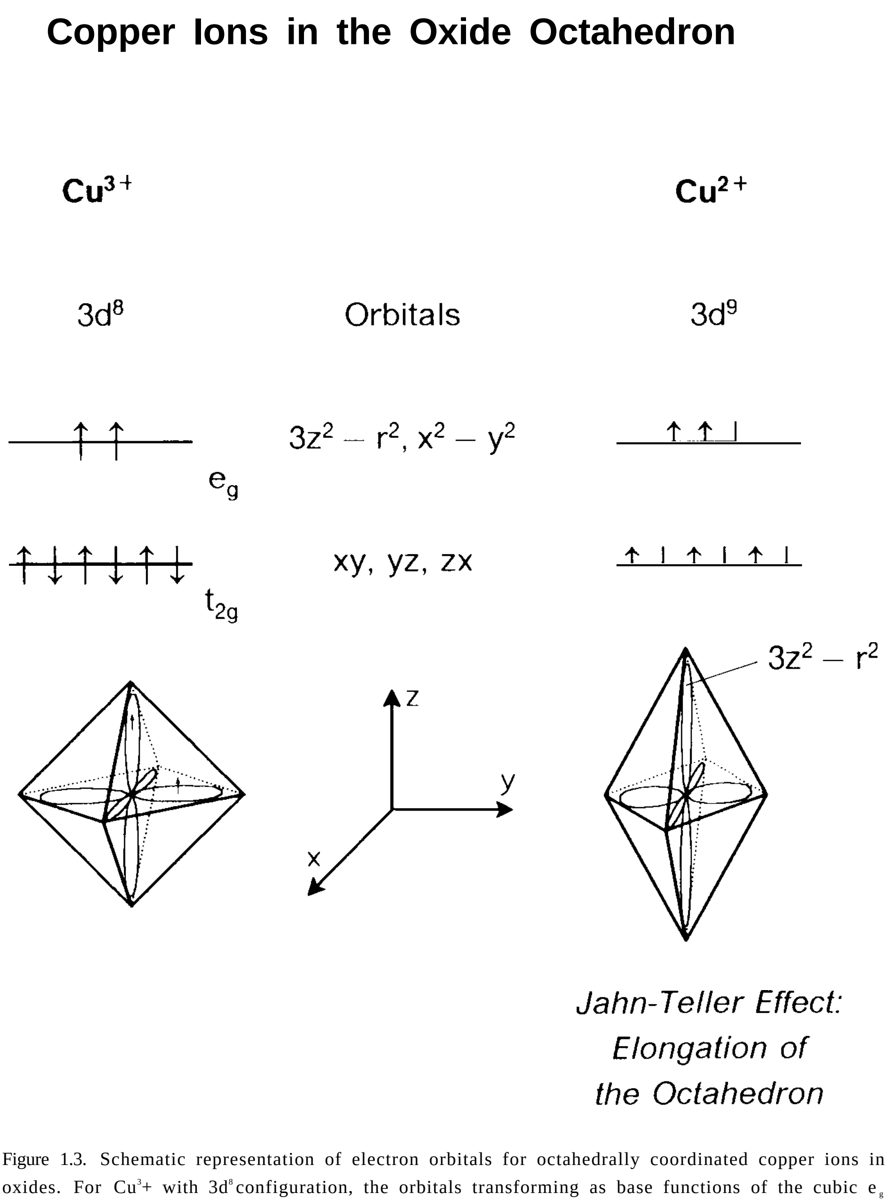


Figure 1.3. Schematic representation of electron orbitals for octahedrally coordinated copper ions in oxides. For Cu^{3+} with $3d^8$ configuration, the orbitals transforming as base functions of the cubic e_g group are half-filled, thus a singlet ground state is formed. In the presence of Cu^{2+} with $3d^9$ configuration? the ground state is degenerate, and a spontaneous distortion of the octahedron occurs to remove this degeneracy. This is known as the Jahn-Teller effect.

also apply to the oxides, our field of expertise, if they could be turned into conductors. We knew there were many of them. Oxides containing TM ions with partially filled e_g orbitals, like Ni^{3+}, Fe^{4+} or Cu^{2+} exhibit a strong JT effect, Figure 1.3, and we considered these as possible candidates for new superconductors.

The Search and Breakthrough

We started the search for high-T, superconductivity in late summer 1983 with the La-Ni-O system. $LaNiO_3$ is a metallic conductor with the transfer energy of the JT-e_g electrons larger than the JT stabilization energy, and thus the JT distortion of the oxygen octahedra surrounding the Ni^{3+} is suppressed [1.13]. However, already the preparation of the pure compound brought some surprises, as the material obtained by our standard coprecipitation method [1.14] and subsequent solid-state reaction turned out to be sensitive not only to the chemicals involved [1.15] but also to the reaction temperatures. Having overcome all difficulties with the pure compound, we started to partially substitute the trivalent Ni by trivalent Al to reduce the metallic bandwidth of the Ni ions and make it comparable to the Ni^{3+} Jahn-Teller stabilization energy. With increasing Al concentration, the metallic characteristics (see Figure 1.4) of the pure $LaNiO_3$ gradually changed, first giving a general increase in the resistivity and finally with high substitution leading to a semiconducting behavior with a transition to localization at low temperatures. The idea did not seem to work out the way we had thought, so we considered the introduction of some internal strain within the $LaNiO_3$ lattice to reduce the bandwidth. This we realized by replacing the La^{3+} ion by the smaller Y^{3+} ion, keeping the Ni site unaffected. The resistance behavior changed in a way we had already recorded in the previous case, and at that point we started wondering whether the target at which we were aiming really did exist. Would the path we decided to embark upon finally lead into a blind alley?

It was in 1985 that the project entered this critical phase, and it probably only survived because the experimental situation, which had generally hampered our efforts, was improved, The period of sharing another group's equipment for resistivity measurements came to an end as our colleague, Pierre Guéret, agreed to my established right to use a newly set-up automatic system. Thus, the measuring time was transferred from late evening to normal working hours. Toni Schneider, at that time acting manager of the Physics department, supported the plans to improve the obsolete x-ray analytical equipment to simplify systematic phase analysis, and in addition, we had some hopes in our new idea, involving another TM element encountered in our search, namely, copper. In a new series of compounds, partial replacement of the JT Ni^{3+} by the non-JT Cu^{3+} increased the absolute value of the resistance, however the metallic character of the solid solutions was preserved down to 4 K [1.13]. But again, we observed no indication of superconductivity. The time to study the literature and reflect on the past had arrived.

It was in late 1985 that the turning point was reached. I became aware of an article by the French scientists C. Michel, L. Er-Rakho and B. Raveau, who

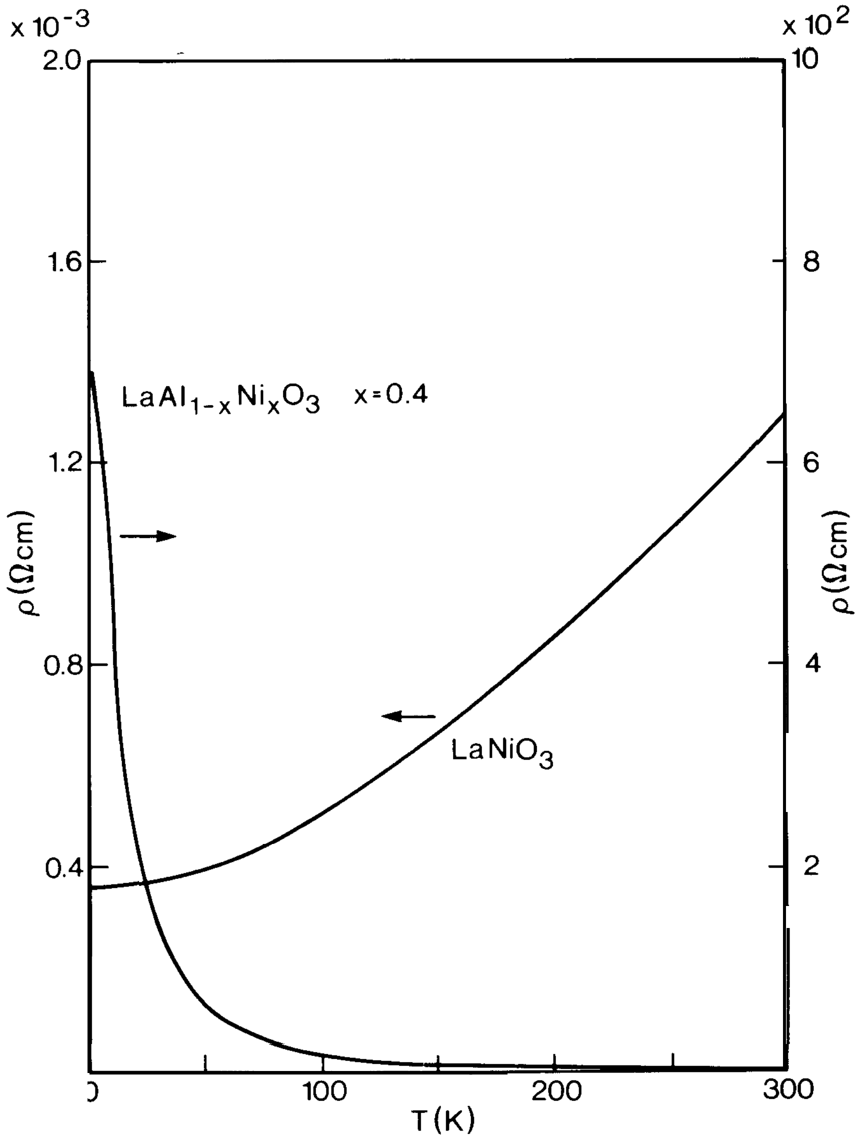


Figure 1.4. Temperature dependence of the resistivity for metallic $LaNiO_3$ and $LaAl_{1-x}Ni_xO_3$, where substitution of Ni^{3+} by Al^{3+} leads to insulating behavior for x = 0.4.

had investigated a Ba-La-Cu oxide with perovskite structure exhibiting metallic conductivity in the temperature range between 300 and - 100° C [1.16]. The special interest of that group was the catalytic properties of oxygen-deficient compounds at elevated temperatures [1.17]. In the Ba-La-Cu oxide with a perovskite-type structure containing Cu in two different valencies, all our concept requirements seemed to be fulfilled.

I immediately decided to proceed to the ground-floor laboratory and start preparations for a series of solid solutions, as by varying the Ba/La ratio one would have a sensitive tool to continuously tune the mixed valency of copper.

Within one day, the synthesis had been performed, but the measurement had to be postponed, owing to the announcement of the visit of Dr. Ralph Gomory, our Director of Research. These visits always kept people occupied for a while, preparing their presentations.

Having lived through this important visit and returning from an extended vacation in mid-January 1986, I recalled that when reading about the Ba-La-Cu oxide, it had intuitively attracted my attention. I decided to restart my activities in measuring the new compound. When performing the four-point

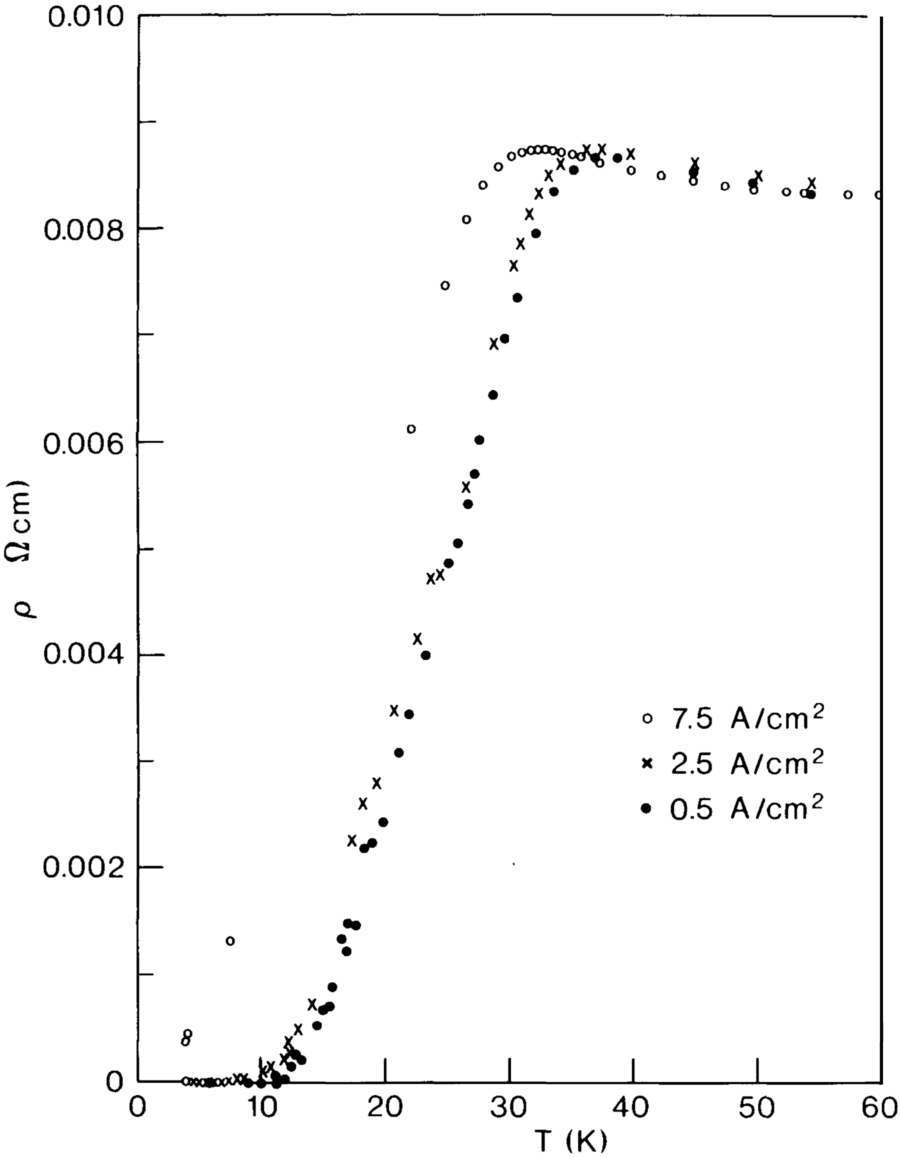


Figure 1.5. Low-temperature resistivity of a sample with x(Ba) = 0.75, recorded for different current densities. From [1.19], © Springer-Verlag 1986.

resistivity measurement, the temperature dependence did not seem to be anything special when compared with the dozens of samples measured earlier. During cooling, however, a metallic-like decrease was first observed, followed by an increase at low temperatures, indicating a transition to localization. My inner tension, always increasing as the temperature approached the 30 K range, started to be released when a sudden resistivity drop of 50% occurred at 11 K. Was this the first indication of superconductivity?

Alex and I were really excited, as repeated measurements showed perfect reproducibility and an error could be excluded. Compositions as well as the thermal treatment were varied and within two weeks we were able to shift the onset of the resistivity drop to 35 K, Figure 1.5. This was an incredibly high value compared to the highest T_c in the Nb_3Ge superconductor.

We knew that in the past there had been numerous reports on high-T, superconductivity which had turned out to be irreproducible [1.7], therefore prior to the publication of our results, we asked ourselves critical questions about its origin. A metal-to-metal transition, for example, was unlikely, owing to the fact that with increasing measuring current the onset of the resistivity drop was shifted to lower temperatures. On the contrary, this behavior supported our interpretation that the drop in $\varrho(T)$ was related to the onset of

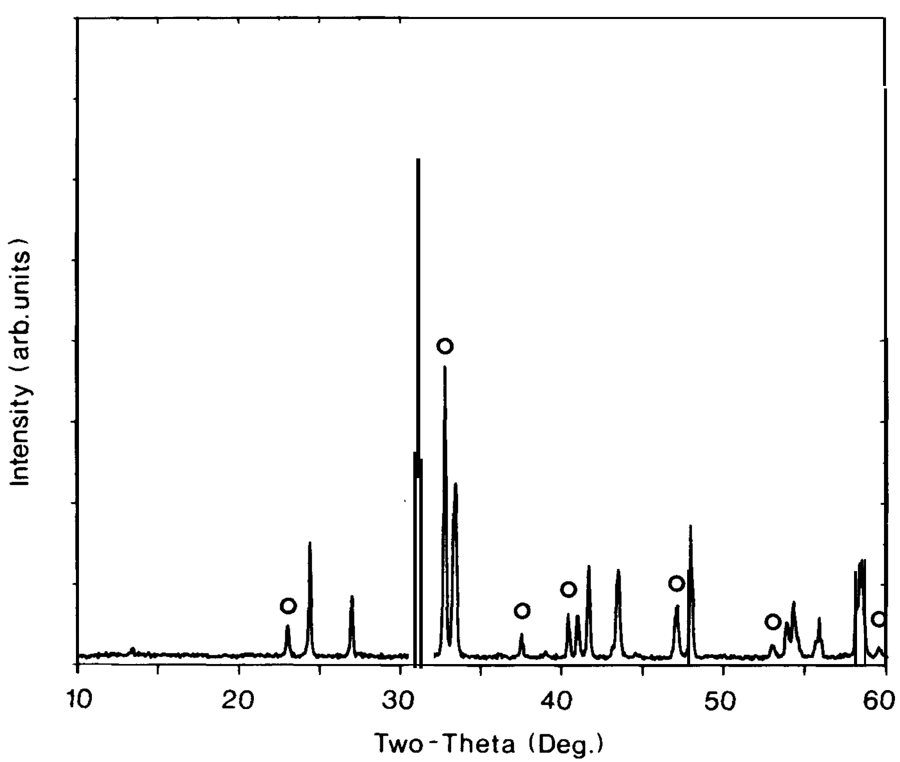


Figure 1.6. X-ray diffraction pattern of a two-phase sample with Ba:La = 0.08. The second phase occurring together with the K_2NeF_4-type phase is indicated by open circles. From [1.20], © 1987 Pergamon Journals Ltd.

superconductivity in granular materials. These are, for example, polycrystalline films of $BaPb_{1-x}Bi_xO_3$ [1.18] exhibiting grain boundaries or different crystallographic phases with interpenetrating grains as in the Li-Ti oxide [1.7]. Indeed, x-ray diffraction patterns of our samples revealed the presence of at least two different phases (see Figure 1.6). Although we started the preparation process of the material with the same cation ratios as the French group, the wet-chemical process did not lead to the same result. This later turned out to be a stroke of luck, in the sense that the compound we wanted to form was not superconducting. The dominating phase could be identified as having a layered perovskite-like structure of K_2NiF_4-type as seen from Figure 1.7. The diffraction lines of the second phase resembled that of an oxygen-deficient perovskite with a three-dimensionally connected octahedra network. In both structures, La was partially replaced by Ba, as we learned from an electron microprobe analysis which Dr. Jürg Sommerauer at the ETH Zurich performed for us as a favor. However, the question was “which is the compound where the mixed valency of the copper leads to the superconductive transition?”

We had difficulties in finding a conclusive answer at the time; however, we rated the importance of our discovery so high that we decided to publish our findings, despite the fact that we had not yet been able to perform magnetic measurements to show the presence of the Meissner-Ochsenfeld effect. Thus, our report was cautiously entitled “Possible High T_c Superconductivity in the Ba-La-Cu-O System” [1.19]. We approached Eric Courtens, my manager at the time, who in late 1985 had already strongly supported our request to purchase a DC Squid Magnetometer, and who is on the editorial board of Zeitschrift für Physik. In this capacity, we solicited his help to receive and submit the paper, although, admittedly, it did involve some gentle persuasion on our part!

Alex and I then decided to ask Dr. Masaaki Takashige whether he would be interested in our project. Dr. Takashige, a visiting scientist from Japan, had joined our Laboratory in February 1986 for one year. He was attached to Alex’s Fellowship group, and I had given him some support in pursuing his activities in the field of amorphous oxides. As he was sharing my office, I was able to judge his reaction, and realized how his careful comments of skepticism changed to supporting conviction while we were discussing the results. We had found our first companion.

Following this, while awaiting delivery of the magnetometer, we tried hard to identify the superconducting phase by systematically changing the composition and measuring the lattice parameters and electrical properties. We found strong indications that the Ba-containing La_2CuO_4 was the phase responsible for the superconducting transition in our samples. Starting from the orthorhombically distorted host lattice, increasing the Ba substitution led to a continuous variation of the lattice towards a tetragonal unit cell [1.20], see Figure 1.8. The highest T_c’s were obtained with a Ba concentration close to this transition (Figure 1.9), whereas when the perovskite phase became dominant, the transition was suppressed and the samples showed only metallic characteristics.

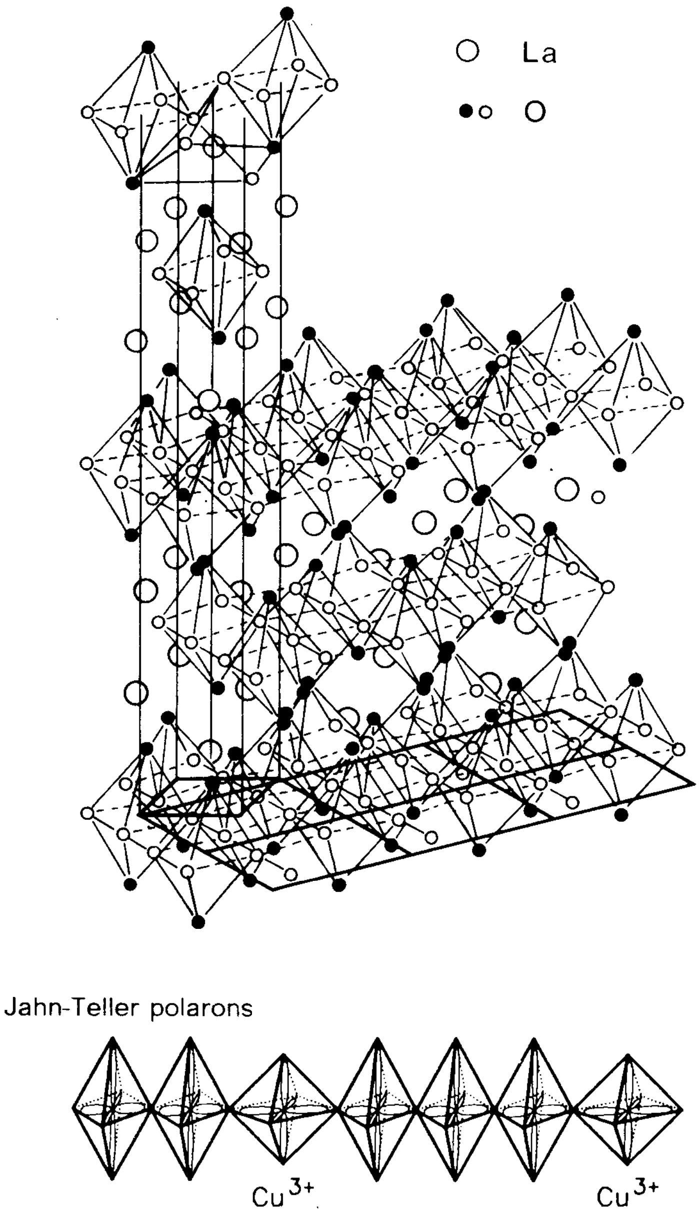


Figure 1.7. Structure of the orthorhombic La_2CuO_4 Large open circles represent the lanthanum atoms, small open and filled circles the oxygen atoms. The copper atoms (not shown) are centered on the oxygen octahedra. From [1.29], © 1987 by the American Association for the Advancement of Science.

The lower part shows schematically how in a linear chain substitution of trivalent La by a divalent alkaline-earth elemcnt would lead to a symmetric change of the oxygen polyhedra in the presence of Cu^{3+}.

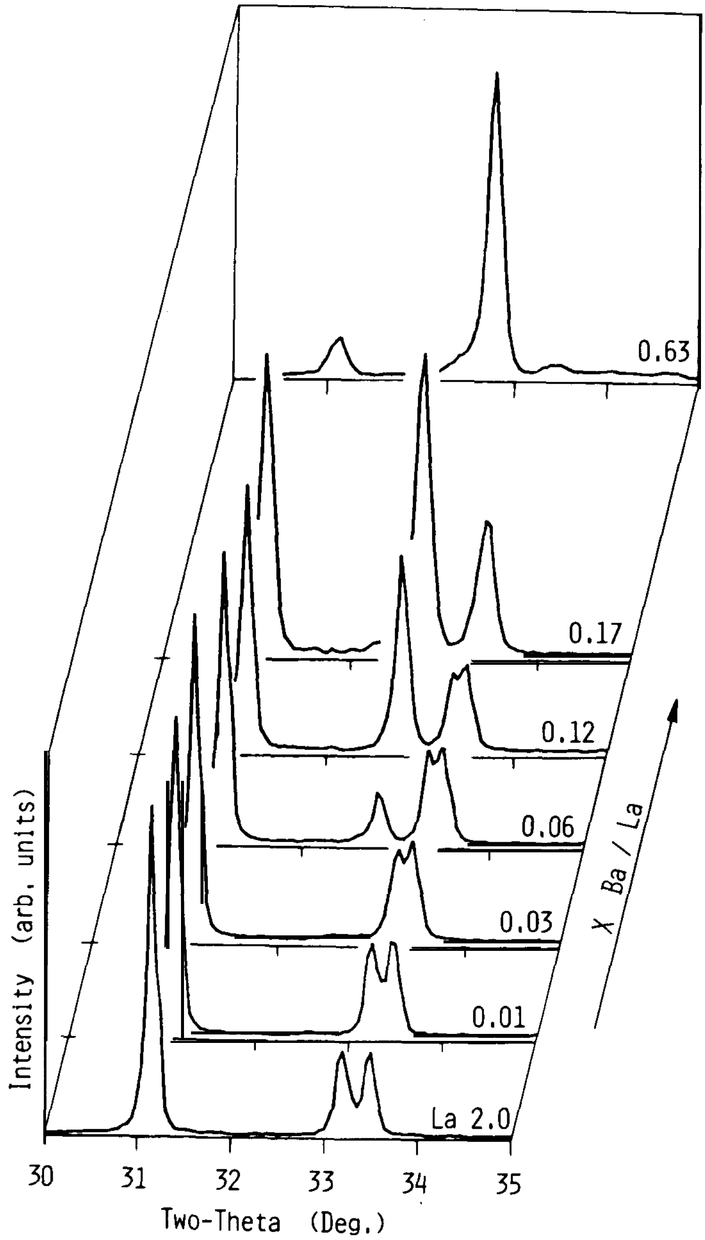


Figure 1.8. Characteristic part of the x-ray diffraction pattern, showing the orthorhombic-to-tetragonal structural phase transition with increasing Ba:La ratio. Concentration axis not to scale. From [1.20], © 1987 Pergamon Journals Ltd.

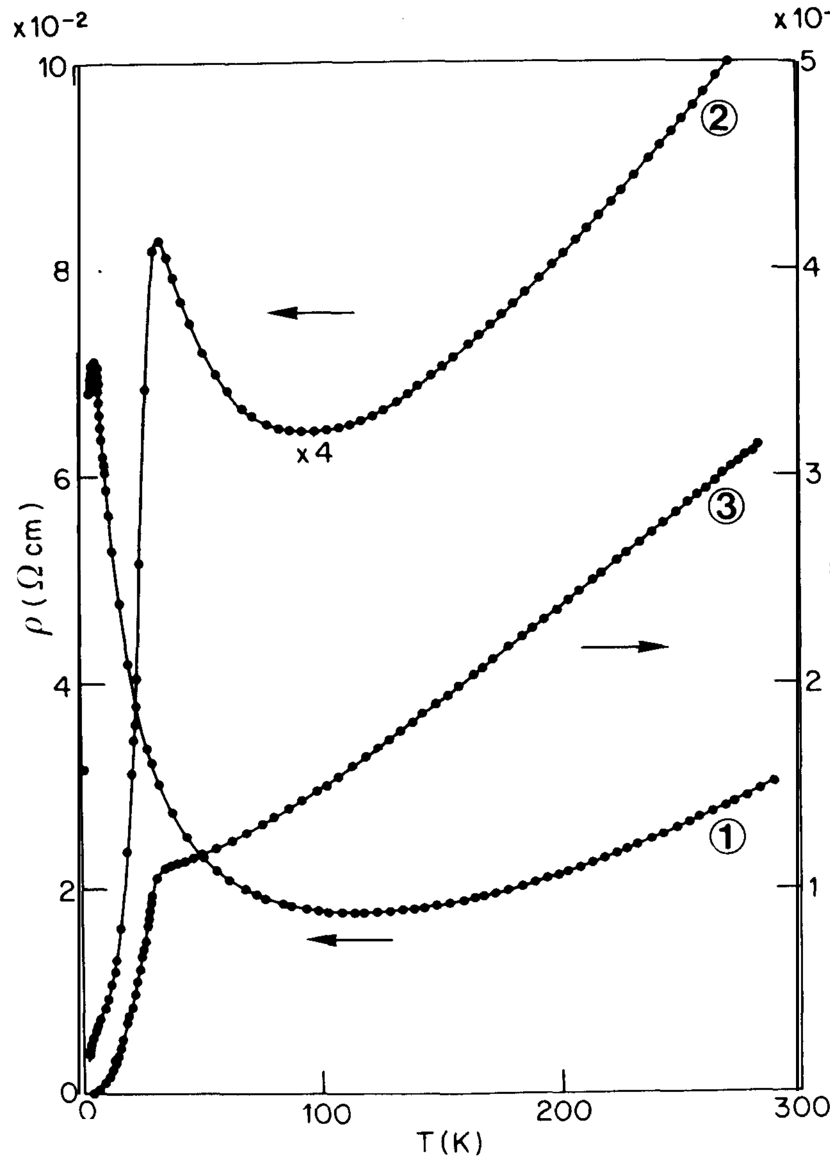


Figure 1.9. Resistivity as a function of temperature for La_2CuO_{4-y}: Ba samples with three different Ba:La ratios. Curves 1, 2, and 3 correspond to ratios of 0.03, 0.06, and 0.07, respectively. From [1.20], © 1987 Pergamon Journals Ltd.

Finally, in September 1986, the susceptometer had been set up and we were all ready to run the magnetic measurements. To ensure that with the new magnetometer we did not measure any false results, Masaaki and I decided to gain experience on a known superconductor like lead rather than starting on our samples. The Ba-La-Cu oxide we measured first had a low Ba content, where metallic behavior had been measured down to 100 K and a transition to localization occurred at lower temperatures. Accordingly, the magnetic susceptibility exhibited Pauli-like positive, temperature-independent and Curie-Weiss behavior at low temperatures, as illustrated by Figure 1.10. Most importantly, within samples showing a resistivity drop, a transition from para- to diamagnetism occurred at slightly- lower temperatures, see Figure 1.11, indicating that superconductivity-related shielding currents existed. The diamagnetic transition started below what is presumably the highest I', in the samples as indicated by theories [1.21, 1.22] describing the behavior of percolative superconductors. In all our samples, the transition to the diamagnetic state was systematically related to the results of our resistivity measurements. The final proof of superconductivity, the presence of the Meissner-Ochsenfeld effect, had

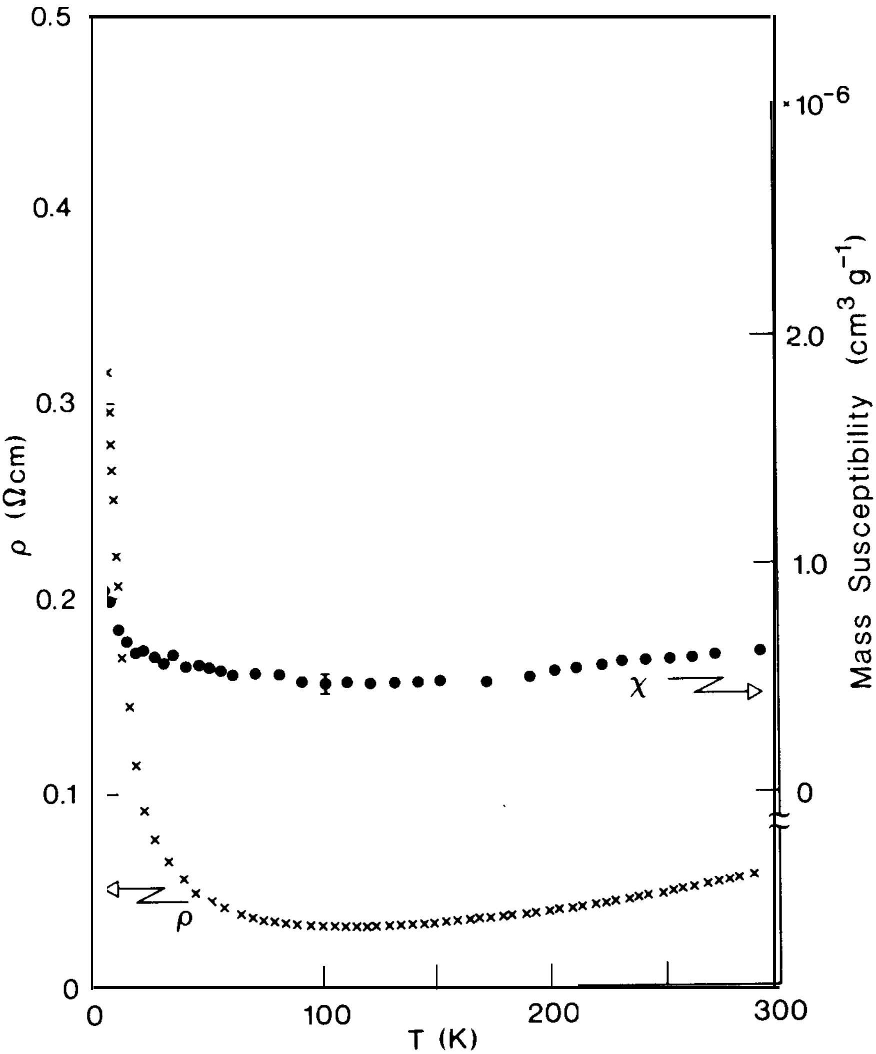


Figure *1.10.* Temperature dependence of resistivity (x) and mass susceptibility (•) of sample 1. From [1.23], ©Les Editions Editions de Physique 1987.

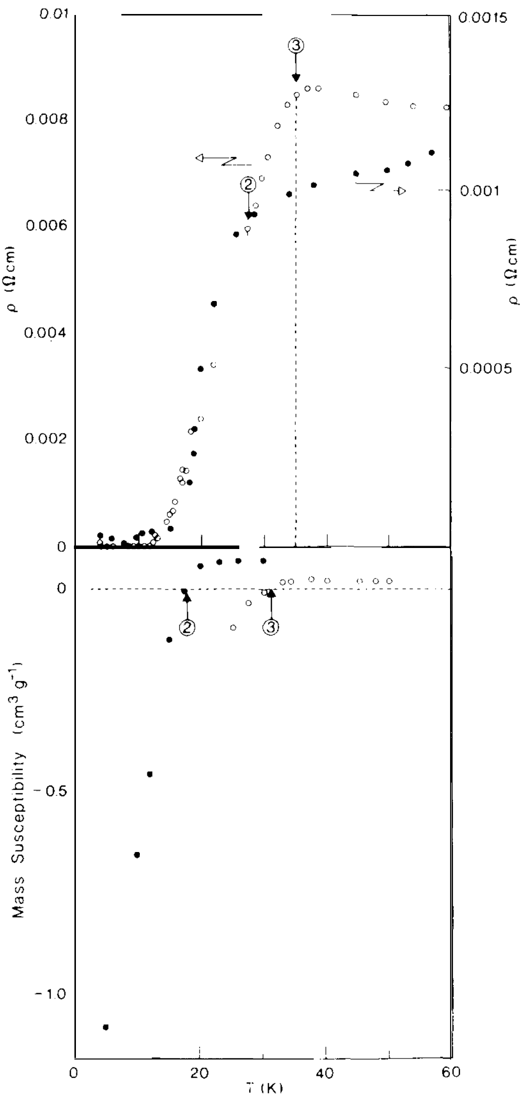


Figure 1.11. Low-tempcrature resistivity and susceptibility of (La-Ba)-Cu-O samples 2 (●) end 3 (○) from [1.23]. Arrows indicate the onset of the resistivity and the paramagnctic-to-diamagnetic transition, respectively. From [1.23], ©Lcs Editions de Physiqe 1987.

been demonstrated. Combining the x-ray analysis, resistivity and susceptibility measurements, it was now possible to clearly identify the Ba-doped La_2CuO_4 as the superconducting compound.

First Responses and Confimations

The number of our troops was indeed growing. Richard Greene at our Research Center in Yorktown Heights had learned about our results and became excited. He had made substantial contributions in the field of organic superconductors and wanted to collaborate in measuring specific-heat data on our samples. We initiated an exchange of information, telefaxing the latest results of our research and sending samples. Realizing that our first paper had appeared in the open literature, we rushed to get the results of our susceptibility data written up for publication.

The day we made the final corrections to our report turned out to be one of the most remarkable days in the history of our Laboratory. Alex, Masaaki and I were sitting together, when the announcement was made over our P.A. system that the 1986 Nobel Prize for Physics had been awarded to our colleagues Gerd Binnig and Heinrich Rohrer. With everything prepared for the submission of our paper [1.23], for one more day we could forget about our work, and together with the whole Laboratory celebrate the new laureates. The next day we were back to reality, and I started to prepare a set of samples for Richard Greene. Praveen Chaudhari, our director of Physical Science in Yorktown Heights, took them with him the same evening

Later in November, we received the first response to our latest work from Professor W. Buckel, to whom Alex had send a preprint with the results of the magnetic measurements. His congratulations on our work were an encouragement, as we began to realize that we would probably have a difficult time getting our results accepted. Indeed, Alex and I had started giving talks about our discovery and, although the presence of the Meissner effect should have convinced people, at first WC were met by a skeptical audience. However, this period turned out be very short indeed.

We continued with the magnetic characterization of the superconducting samples and found interesting properties related to the behavior of a spin glass [1.24]. We then intensively studied the magnetic field and time dependences of the magnetization, before finally starting to realize an obvious idea, namely, to replace La also by other alkaline-earth elements like Sr and Ca. Especially Sr^{2+} had the same ionic radius as La^{3+}. We began experiments on the new materials which indicated that for the Sr-substituted samples 'I', was approaching 40 K and the diamagnetism was even higher, see Figure 1.12(a) and (b), [1.25]. It was just at that time that we learned from the Asahi-Shinbun International Satellite Edition [November 28, 1986) that the group of Professor Tanaka at the University of Tokyo had repeated our experiments and could confirm our result [1.26]. We were relieved, and even more so when we received a letter from Professor C. W. Chu at the University of Houston, who was also convinced that within the Ba-La-Cu-O system superconductivity occurred at 35 K [1.27]. Colleagues who had not paid any attention to our work at all suddenly became alert. By applying hydrostatic pressure to the samples, Professor Chu was able to shift the superconductive transition from 35 to almost 50 K [1.27]. Modification of the original oxides by introducing the smaller Y^{3+}, for the larger La^{3+} resulted in a giant jump of T_c to 92 K in multiphase samples [1.28], Figure 1.13 At a breathtaking pace, dozens of groups now repeated these experiments, and after an effort of only a few days the new superconducting compound could be isolated and identified. The resistive transition in the new $YBa_2Cu_3O_7$ compound was complete at 92 K, Figure 1.14, and even more impressive was the fact that the Meissner effect

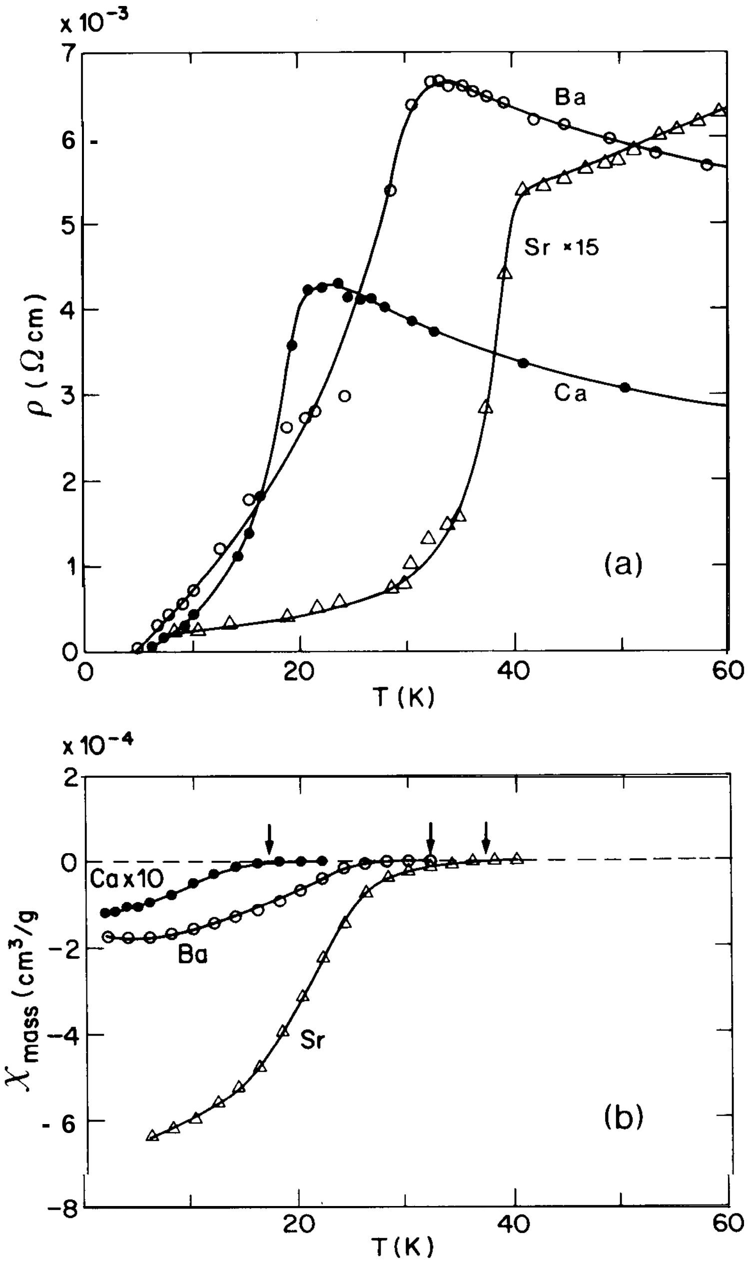


Figure 1.12. (a) Resistivity as a function of temperature for Ca (●), Sr (△), and Ba (○) substitution with substituent-to-La ratios of 0.2/1.8, # 0.2/1.8, 0.15/1.85, respectively. The Sr curve has been vertically expanded by a factor of 15. (b) Magnetic susceptibility of these samples. The substituents are Ca (●), Sr (△), and Ba (○), with total sample masses of 0.14, 0.21, and 0.13 g, respectively. The Ca curve has been expanded by a factor of 10. Arrows indicate onset temperatures. From [1.25], © 1987 by the American Association for the Advancement of Science.

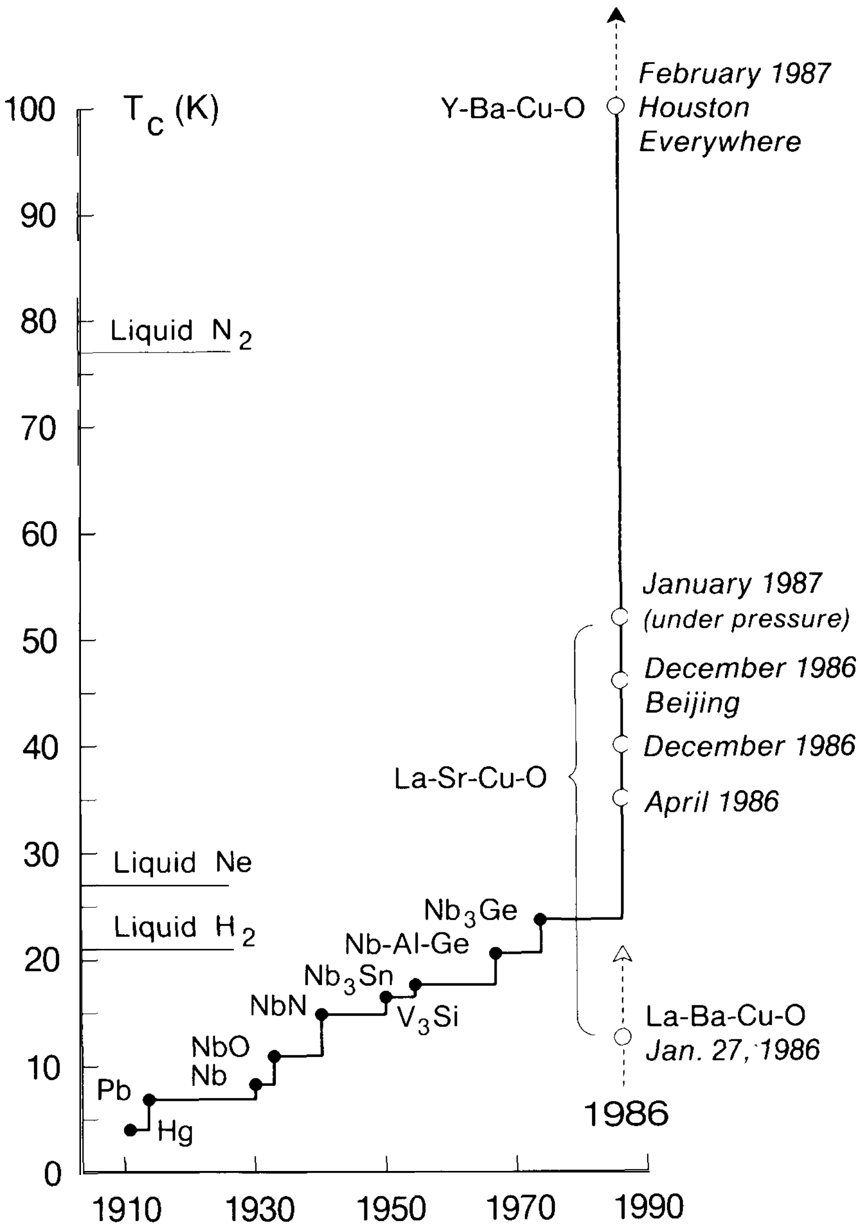


Figure 1.13. Evolution of the superconductive transition temperature subsequent to the discovery of the phenomenon. From [1.29], © 1987 by the American Association for the Advancement of Science.

could now be demonstrated without any experimental difficulties with liquid nitrogen as the coolant. Within a few months, the field of superconductivity had experienced a tremendous revival, with an explosive development of T_c's which nobody can predict where it will end.

An early account of the discovery appeared in the September 4, 1987, issue of *Science,* which was dedicated to science in Europe [1.29].

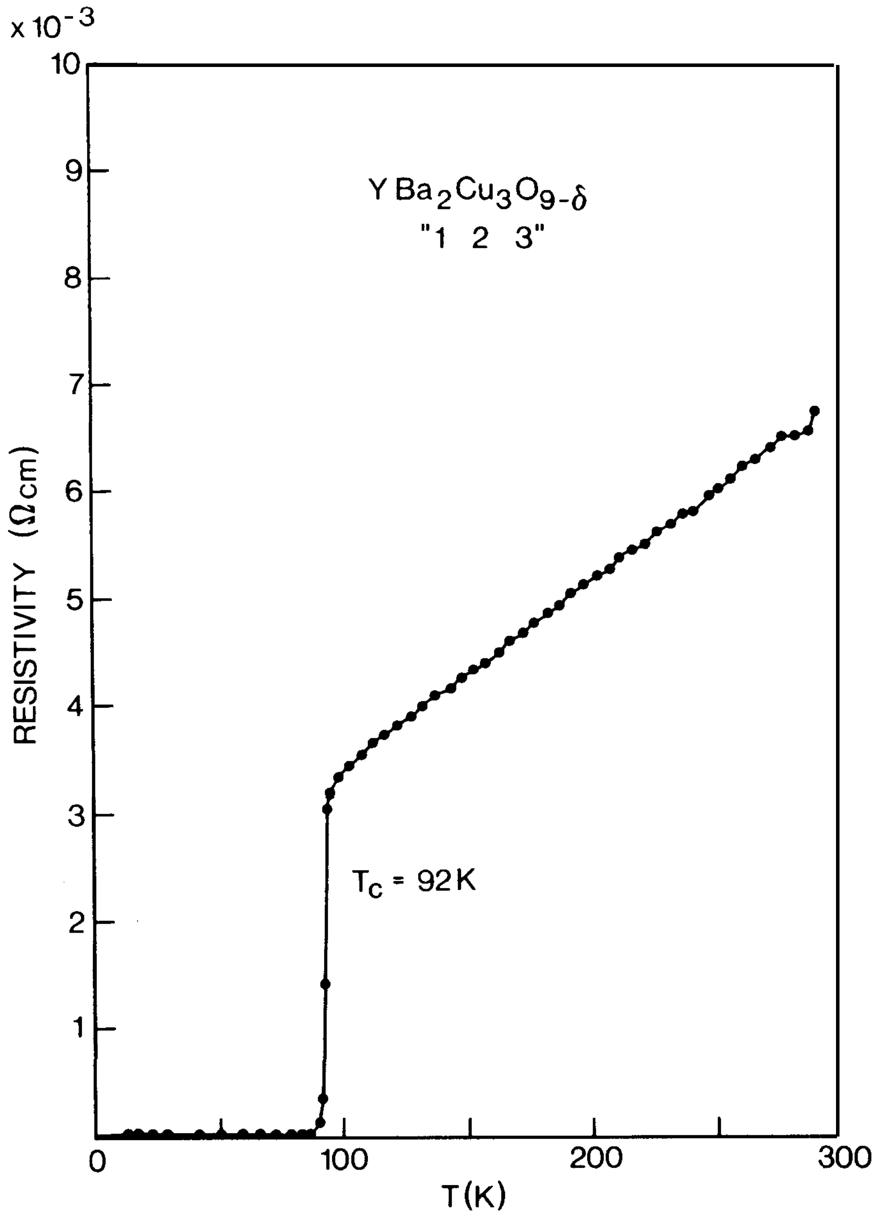


Figure 1.14. Resistivity of a single-phase $YBa_2Cu_3O_7$ sample as a function of temperature.

References Part 1

[1.1] Müller, K. A. (1971) in: Structural *Phase Transitions and Soft Modes,* editors E. J. Samuelsen, E. Andersen, and J. Feder (Universitetsforlaget, Oslo) p. 85.

[1.2] Schneider, T. and Stoll, E. (1971) Physica 55, 702.

[1.3] Schooley, J. F., Frederikse, H. P. R., Hosler, W. R., and Pfeiffer, E. R. (1967) Phys. Rev. *159,* 301.

[1.4] Baratoff, A. and Binnig, G. (1981) Physica *108B,* 1335; Baratoff, A., Binnig, G., Bednorz, J. G., Gervais, F., and Servoin, J. L. (1982) in: *Superconductivity in d- and f-Band Metals, Proceedings IV Conference on Superconductivily in d- and f-Band Metals,* editors W. Buckel and W. Weber (Kernforschungszentrum Karlsruhe) p. 419.

[1.5] Binnig, G., Baratoff, A., Hoenig, H. E., and Bednorz, J. G. (1980) Phys. Rev. Lett. 45, 1352.

[1.6] Müller, K. A., Pomerantz, M., Knoedler, C. M., and Abraham, D. (1980) Phys. Rev. Lett. 45, 832, and references therein.

[1.7] Beasley, M. R. and Geballe, T. H. (1984) Phys. Today 36(l0), 60; Muller, J. (1980) Rep. Prog. Phys. 43, 663.

[1.8] Johnston, D. C., Prakash, H., Zachariasen, W. H., and Viswanathan, R. (1973) Mat. Res. Bull. 8, 777.

[1.9] Sleight, A. W., Gillson, J. L., and Bierstedt, P. E. (1975) Solid State Commun. *17, 27.*

[1.10] Bardeen, J., Cooper, L. N., and Schrieffer, J. R. (1957) Phys. Rev. 108, 1175.

[1.11] Chakraverty, B. K. (1979) J. Physique Lett. 40, L99; *idem* (1981) J. Physique 42, 1351.

[1.12] Höck, K.-H., Nickisch, H., and Thomas, H. (1983) Helv. Phys. Acta 56, 237.

[1.13] Goodenough, J. B. and Longo, M. (1970, "Magnetic and other properties of oxide and related compounds" In: *Landoldt-Boernstein New Series. Vol. III/4a: Crystal and Solid State Physics,* editors K. H. Hellwege and A. M. Hellwege (Springer-Verlag, Berlin, Heidelberg, New York), p. 262, Fig. 73.

[1.14] Bednorz, J. G., Müller, K. A., Arend, H., and Gränicher, H. (1983) Mat. Res. Bull. *18,* 181.

[1.15] Vasanthacharya, N. Y., Ganguly, P., Goodenough, J. B., and Rao, C. N. R. (1984) J. Phys. C: Solid State Phys. *17, 2745.*

[1.16] Michel, C., Er-Rakho, L., and Raveau, B. (1985) Mat. Res. Bull. 20, 667.

[1.17] Michel, C. and Raveau, B. (1984) Rev. Chim. Min. *21,* 407. It is ofinterest that back in 1973, Goodenough, J. B., Demazeaux, G., Pouchard, M., and Hagenmüller, P. (1973) J. Solid State Chem. 8, 325, in Bordeaux, and later Shaplygin, I. S., Kakhan, B. G., and Lazarev, V. B. (1979) Russ. J. Phys. Chem. 24(6), 1478, pursued research on layered cuprates with catalysis applications in view.

[1.18] Suzuki, M., Murakami, T., and Inamura, T. (1981) Shinku 24, 67; Enomoto, Y., Suzuki, M., Murakami, T., Inukai, T., and Inamura, T. (1981) Jpn. J. Appl. Phys. 20, L661.

[1.19] Bednorz, J. G. and Müller, K. A. (1986) 'Z. Phys. B 64, 189.

[1.20] Bednorz, J. G., Takashige, M., and Müller K. A. (1987) Mat. Res. Bull. 22, 819.

[1.21] Bowman, D. R. and Stroud, D. (1984) Phys. Rev. Lett. 52, 299.

[1.22] Ebner, C. and Stroud, D. (1985) Phys. Rev. B 31, 165, and references therein.

[1.23] Bednorz, J. G., Takashige: M., and Müller, K. A. (1987) Europhys. Lett. 3. 379.

[1.24] Müller, K. A., Takashige, M., and Bednorz, J. G. (1987) Phys. Rev. Lett. 58, 1143.

[1.25] Bednorz, J. G., Müller, K. A.. and Takashige, M. (1987) Science 236, 73.

[1.26] Takagi, H., Uchida, S., Kitazawa, K., and Tanaka, S. (1987) Jpn. J. Appl. Phys. 26, L123; Uchida, S., Takagi, H., Kitazawa, K., and Tanaka, S. (1987) *ibid.,* L151.

[1.27] Chu, C. W., Hor, P. H., Meng, R. L., Gao, I,., Huang, Z. J., and Wang, Y. Q. (1987) Phys. Rev. Lett. 58, 405.
[1.28] Wu, M. K., Ashburn, J. R., Torng, C. J., Hor, P. H., Meng, R. L., Goa, I.., Huang, Z. J., Wang, Y. Q., and Chu, C. W. (1987) Phys. Rev. Lett. 58. 908; Hor, P. H., Gao, L., Meng, R. L., Huang, Z. J., Wang, Y. Q., Forster, K., Vassilious, J., Chu, C. W., Wu, M. K., Ashburn, J. R., and Torng, C. J. (1987) ibid., 91 1.
[1.29] Müller, K. A. and Bednorz, J, G. (1987) Science 237, 1133.

PART 2: PROPERTIES OF THE NEW SUPERCONDUCTORS

In the second part, properties of the new layered oxygen superconductors were described. Since their discovery, summarized in the first part, a real avalanche of papers has been encountered; thus it would be beyond the scope of this Lecture to review all of them here. A forthcoming international conference in Interlaken, Switzerland, in February 1988, is intended to fulfill this task and will be chaired by one of us. Therefore, only a selected number of experiments were presented in Stockholm; those judged of importance at this time for the understanding of superconductivity in the layered copper oxides. In some of them, the laureates themselves were involved, in others not. Owing to the frantic activity in the field, it may be possible that equivalent work with priority existed unbeknown to us. Should this indeed be the case, WC apologize and propose that the following be read for what it is, namely, a write-up of the lecture given, including the transparencies shown.

After the existence of the new high-T, superconductors had been confirmed, one of the first questions was "What type of superconductivity is it?" Does one again have Cooper pairing [2.1] or not? This question could be answered in the affirmative. The earliest experiment to come to our knowledge was that of the Saclay-Orsay collaboration. Estève et al. [2.2] measured the I-V. characteristics of sintered $La_{1.85}Sr_{0.15}CuO_4$ ceramics using nonsuperconducting Pt-Rh, Cu or Ag contacts. In doing so, they observed weak-link characteristics internal to the superconductor, to which we shall subsequently revert. Then they applied microwaves at $\varkappa$ = 9.4 GHz and observed Shapiro steps [2.3] at V_s = 19 µV intervals. From the well-known Josephson formula [2.4]

$$V_s = h\varkappa/q, \tag{2.1}$$

they obtained q = 2e, i.e., Cooper pairs were present. Figure 2.1 illustrates these steps. From the fundamental London equations, the flux ϕ through a ring is quantized [2.5]

$$\begin{aligned} \phi &= n\phi_0 \\ \phi_0 &= hc/q \end{aligned} \tag{2.2}$$

The clearest experiment, essentially following the classical experiments in 1961, was carried out in Birmingham, England, by- C. E. Gough el al. [2.6]. They detected the output of an r.f:-SQUID magnetometer showing small integral numbers of flux quantum ϕ_0 jumping into and out of the ring of $Y_{1.2}Ba_{0.8}CuO_4$, see Figure 2.2. The outcome clearly confirmed that q = 2e.

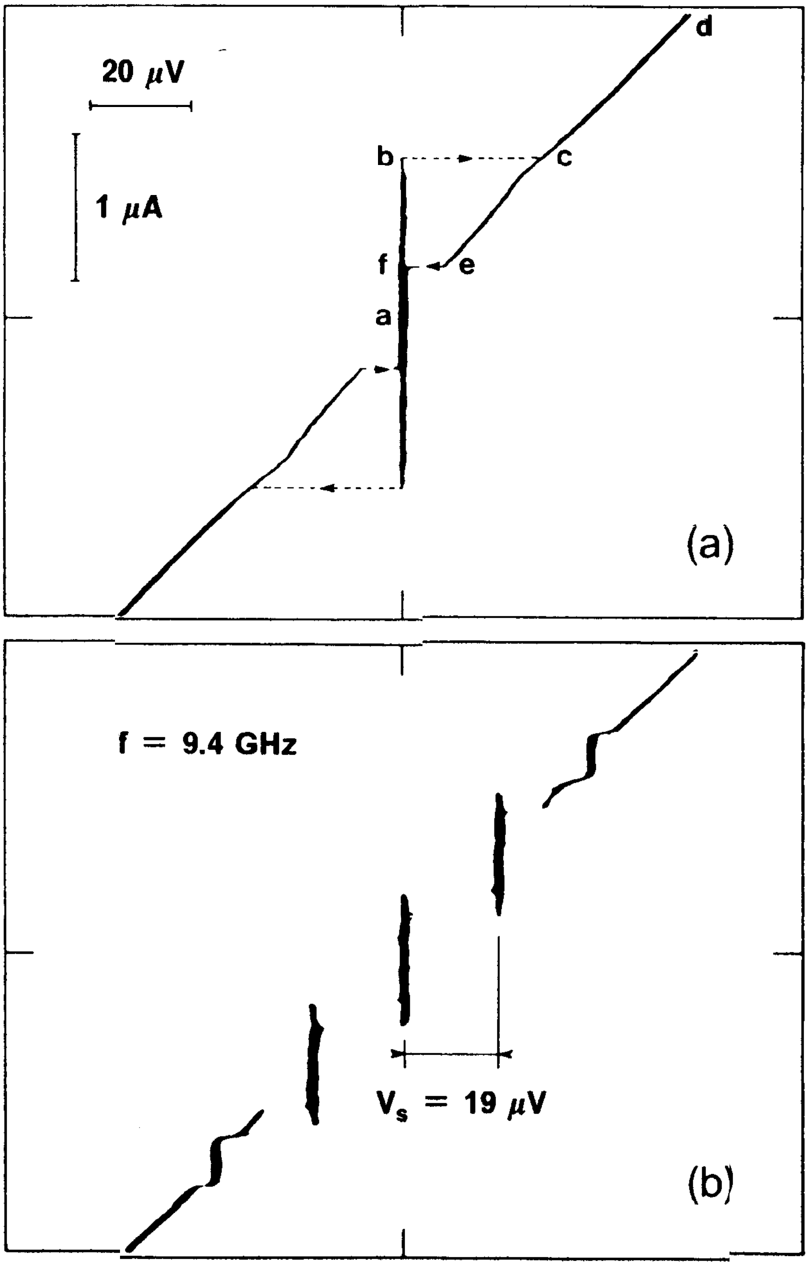


Figure 2.1. (a) Oscilloscope trace of a current-voltage characteristic obtained at 4.2 K with an aluminum tip on a $La_{1.85}Sr_{0.15}CuO_4$ sample. Letters a through f indicate sense of trace. Dashed lines have been added to indicate the switching between the two branches. (b) Steps induced by a microwave irradiation at frequency f = 9.4 GHz. All other experimental conditions identical with those of (a). From [2.2] © Les Editions de Physique 1987.

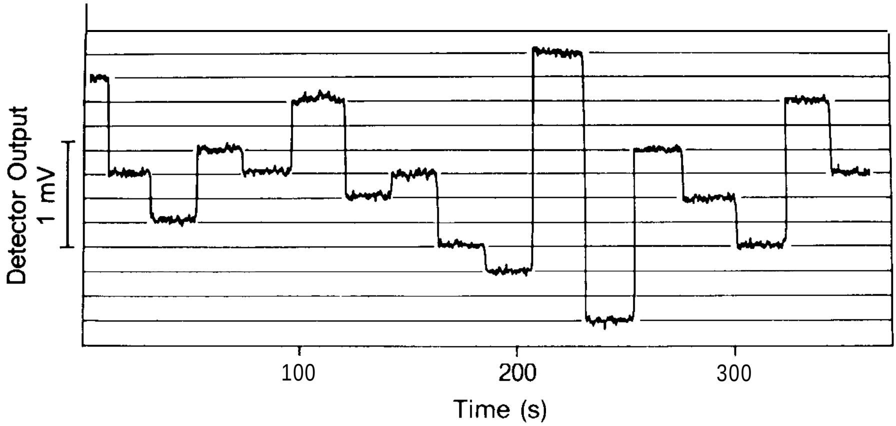


Figure 2.2. Output of the r.f.-SQUID magnetometer showing small integral numbers of flux quanta jumping in to and out of a ring. Reprintcd by permission from [2.6], copyright © 1987 Macmillan Magazines Ltd.

To understand the mechanism, it was of relevance to know the nature of the carrier charge present. In La_2CuO_4 doped very little, the early measurements [1.23] showed localization upon doping with divalent Ba^{2+} or Sr^{2+} and Ca^{2+}; it was most likely that these ions substituted for the trivalent La^{3+} ions. Thus, from charge-neutrality requirements, the compounds had to contain holes. Subsequent thermopower and Hall-effect measurements confirmed this assumption [2.7]. The holes were thought to be localized on the Cu ions. Because the copper valence is two in the stoichiometric insulator La_2CuO_4, doping would create Cu^{3+} ions. Thus a mixed Cu^{2+}/Cu^{3+} state had to be present. By the same argument, this mixed-valence state ought also to occur in YBa_2CuO_7-,(6 ~ 0.1). Early photo-electron core-level spectra (XPS and UPS) by Fujimori el al. [2.8] and Bianconi et al. [2.9] in $(La_{1-x}Sr_x)_2CuO_{4-y}$ and $YBa_2La_3Cu_{6\cdot7}$ did not reveal a $\underline{2p}\,3d^8$ final state owing to a $Cu^{3+}\,3d^8$ state (the underlining indicates a hole). However, the excitation was consistent with the formation of holes L in the oxygen-derived band, i.e., a predominant $3d^9L$ configuration for the formal Cu^{3+} state. Photo-x-ray absorption near the edge structure was also interpreted in the same manner by comparison to other known Cu compounds. Emission spectra by Petroff's group [2.10] pointed in the same direction since the excitation thresholds were compatible with the presence of holes in Cu-O hybrid bands. From their data, both groups concluded that strong correlation effects were present for the valence carriers. However, these results were challenged by other groups working in the field, partially because the spectra involved the interpretation of Cu-atom satellites. A beautifully direct confirmation of the presence of holes on the oxygen p-levels, like L, was carried out by Nücker *et al.* [2.11]. These authors investigated the core-level excitation of oxygen 1s electrons into empty 2p states of oxygen at 528 eV. This is an oxygen-specific experiment. If no holes are present on the p-level, no absorption will occur. Figure 2.3 summarizes their data on $La_{2-x}Sr_xCuO_4$ and $YBa_2Cu_3O_{7-\delta}$. It is shown that for $x \approx 0$ and6 = 0.5, no

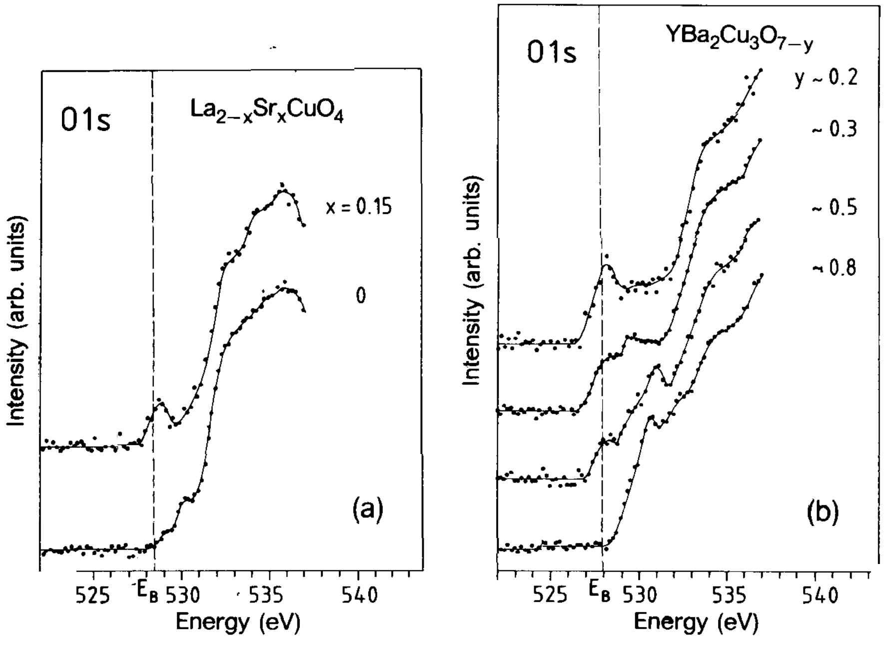


Figure 2.3. Oxygen ls absorption edges of (a) $La_{2-x}Sr_xCuO_4$ and (b) $YBa_2Cu_3O_{7-y}$ measured by energy-loss spectroscopy. The binding energy of the O ls level, as determined by x-ray induced photoemission, is shown by the dashed line. In the framework of an interpretation of the spectra by the density of unoccupied states, this line would correspond to the Fermi energy. From [2.11], © 1988 The American Physical Society.

oxygen p-holes are present and thus no absorption is observed, whereas upon increasing x or reducing δ, a $\underline{2p}$ hole density at the Fermi level is detected in both compounds.

Of substantial interest is the dependence of the transition temperature on the hole concentration. The electron deficiency is hereafter written in the form $[Cu-O]^+$ as a peroxide complex in which the probability of the hole is about 70% $3d^9\,\underline{2p}$, as discussed above, and 30% $3d^8$ as recently inferred from an XPS study [2.12]. Hall-effect data are difficult to analyze in the presence of two-band conductivity, which is possible in these copper-oxide compounds, owing to the well-known compensation effects. Therefore, M. W. Shafer, T. Penney and B. L. Olson [2.13] determined the concentration by wet chemistry according to the reaction $[Cu-O]^+ + Fe^{2+} \leftrightarrows Cu^{2+} + Fe^{3+} + O^{2-}$ in the $La_{2-x}Sr_xCuO_{4-\delta}$ compound. Figure 2.4 shows a plot of T_c vs $[Cu-O]^+$ concentration with a maximum of 35 K of 15% total copper present. There is also a clear threshold at about 5%. From the study, it is apparent that 15-16% $[Cu-0]^+$ is the maximum number of holes the La_2CuO_4 structure accepts. Beyond this concentration, oxygen vacancies are formed. The relationship between T_c and $[Cu-O]^+$ in $La_{2-x}Sr_xCuO_4$ was extended to $YBa_2Cu_3O_{7-\delta}$. The inset of Figure 2.4 illustrates the results under the assumption that two layers in the 123 compound are active for $\delta \approx 0.1$ and $\delta \approx 0.3$, i.e., T_c's of 92 K and 55 K [2.14]. The latter transition, first reported by Tarascon

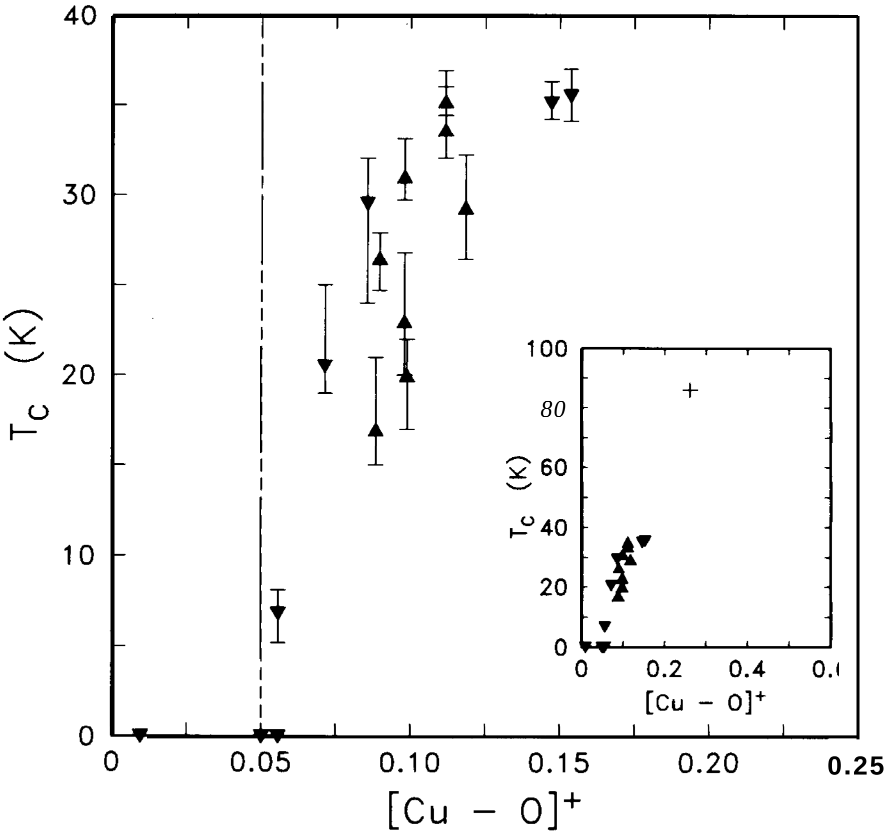


Figure 2.4. T_c vs the hole concentration $[Cu\text{-}O]^+$, as a fraction of total copper. Down triangles are for compositions with x<0.15, up triangles for x>0.15. Inset shows same data plus points for single $YBa_2Cu_3O_{6.6}$ sample as discussed in the text. From [2.13], ©1987 The American Physical Society.

and coworkers, could be well evidenced by near-room-temperature plasma oxidation of the oxygen-deficient Y-compound [2.14].

The $La_{2-x}Sr_xCuO_{4-y}$ with its less complicated structure allows easier testing of models. Its magnetic properties below the hole threshold concentration, x = 0.05, are of special interest. For x = 0, the susceptibility x(T) = M(T)/H exhibits a maximum at low fields of H = 0.05 Tesla below 300 K. This maximum increases in height and shifts to lower temperatures for higher magnetic fields up to 4.5 T [2.15] as seen in Figure 2.5(a). Such behavior is indicative of spin density waves or antiferromagnetic fluctuations. Indeed, neutron-diffraction experiments by Vaknin et al. [2.16] proved three-dimensional (3-D) antiferromagnetic ordering up to 240 K depending on oxygen stoichiometry (i.e., hole concentration). The structure is shown in Figure 2.5(b). Subsequent neutron-scattering experiments on a single crystal revealed a novel two-dimensional (2-D) antiferromagnetic correlation well above and also *below* the 3-D Néel temperature of T_N as shown in Figure 2.6. This instantaneous (not time-averaged) ordering was seen even above room temperature [2.17]. The existence of antiferromagnetism (A.F.) supports models in which holes lead either

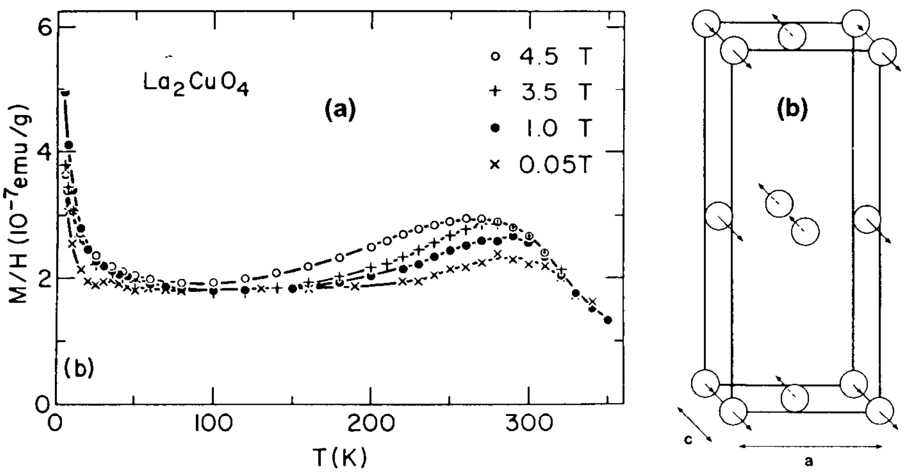


Figure 2.5. (a) Temperature dependence of the magnetic susceptibility $\chi = M/H$ of La_2CuO_4 in different fields H. From [2.15], © 1987 Pergamon Journals Ltd. (b) Spin structure of antiferromagnetic La_2CuO_{4-y}. Only copper sites in the orthorhombic unit cell are shown for clarity. From 12.16], © 1987 The American Physical Society.

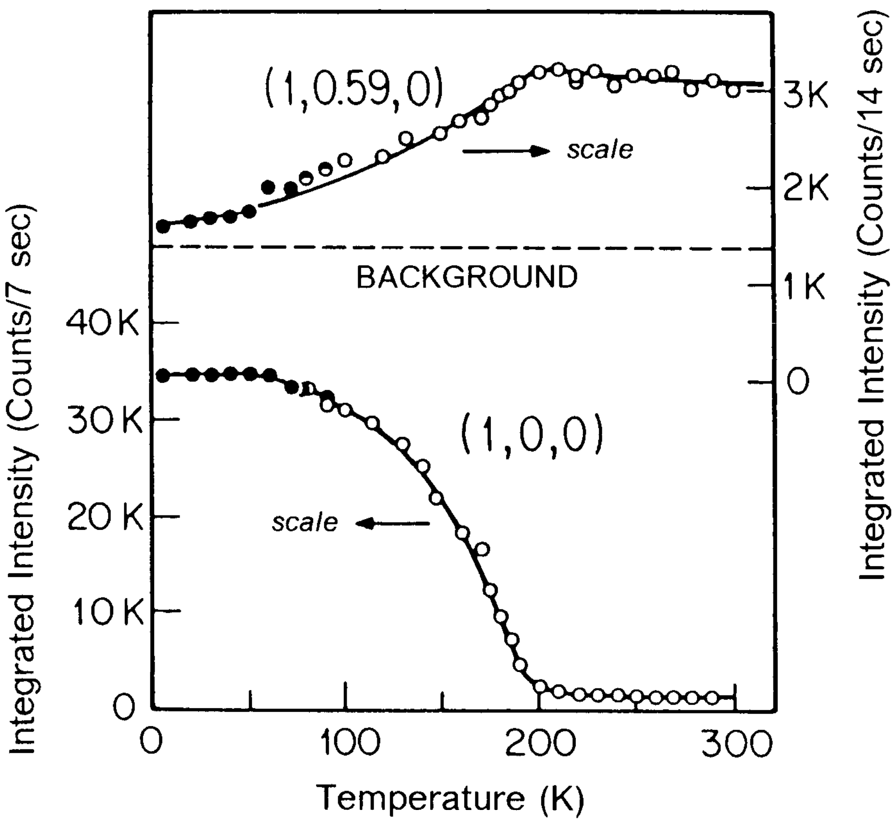


Figure 2.6. Integrated intensities of the (100) 3-D antiferromagnetic Bragg peak and the (1,0.59,0) 2-D quantum spin fluid ridge. The open and filled circles represent separate experiments. From [2.17], © 1987 The American Physical Society.

to localization or to pairing in the strong-coupling limit as proposed by Emery [2.18] and others [2.19]. The resonant valence-band state is also related to the A.F. state [2.20].

From the prevalence of magnetic interactions as primary cause for the occurrence of the high-T, superconductivity, one would expect the isotope effect to be absent. This, because the latter effect is found when the Cooper pairing is mediated by phonon interaction, as found in most of the metallic superconductors previously known. Indeed, substituting O^{16} by O^{18} in the $YBa_2Cu_3O_{7-\delta}$ compound at AT&T did not reveal a shift in T_c [2.21]. However, substitution experiments in the $La_{2-x}Sr_xCuO_{4-y}$ carried out shortly thereafter did reveal an isotope effect with $0.14 < \beta < 0.35$ [2.22] as compared with the full effect of $\beta = 1/2$ deduced from the weak-coupling formula [1. 10]

$$T_c = 1.13\Theta_D \exp-(1/N(E_F)_x V^*), \tag{2.3}$$

with the Debye temperature $\Theta_D \propto 1/M^{1/2}$ of the reduced mass. Thus in the lanthanum compound, oxygen motion is certainly present. As it is highly unlikely that the mechanism is substantially different in the 123 compound, oxygen motion should also be there. This, because absence of the isotope effect does not necessarily exclude a phonon mechanism, which has to be present if Jahn-Teller polarons participate. Indeed, a subsequent, more accurate experiment did show a weak isotope effect in $YBa_2Cu_3O_7$ with AT, $\approx$ 0.3 to 0.5 K [2.23]. From these results, it appears likely that there is more than one interaction present which leads to the high transition temperatures, the low quasi-2D properties certainly being of relevance.

The x-ray and photoemission studies mentioned earlier had indicated strong correlation effects. Cooper pairing having been ascertained, it was therefore of considerable interest whether the new superconductors were of the strong- or the weak-coupling variety. In the latter case, the gap 2A to kT_c, ratio is [1. 10]

$$\frac{2\Delta}{kT_c} = 3.52, \tag{2.4}$$

whereas in the former it is larger.

Tunneling experiments have been widely used to determine the gap in the classical superconductors. However, the very short coherence length yields too low values of 2Δ, as will be discussed later [2.24]. Infrared transmission and reflectivity measurements on powders were carried out at quite an early stage. With the availability of $YBa_2Cu_3O_7$ single crystals, powder infrared data are less relevant, but are quoted in the more recent work. An interesting example is the reflectivity study by Schlesinger et al. [2.25] of superconducting $YBa_2Cu_3O_7$ and a Drude tit to the nonsuperconducting $YBa_2Cu_3O_{6.5}$ data. From the Mattis-Bardeen enhanced peak in the superconducting state, these authors obtained $2\Delta_{ab}/kT_c \approx 8$, i.e., strong coupling in the Cu-O planes, see Figure 2.7. NMR relaxation experiments by Mali *et* al. [2.26], although not yet completely analyzed, yield two gaps with ratios 4.3 and 9.3, respectively i.e., the latter in the range of the infrared data.

NMR relaxation experiments were among the first at the time to prove the

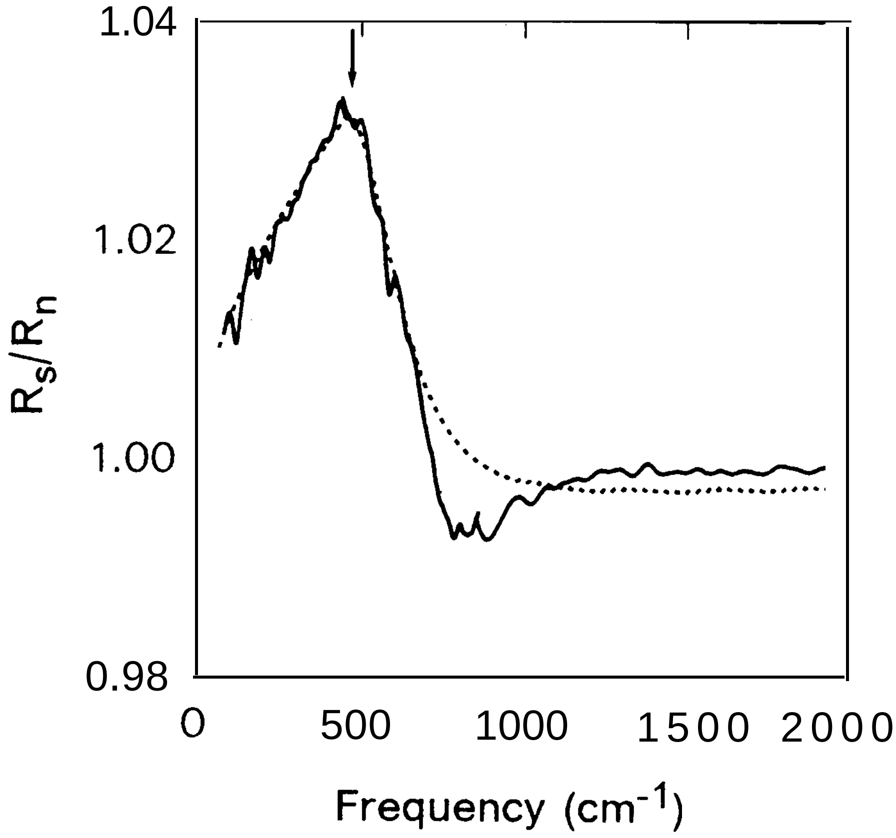


Figure 2.7. Normalized infrared reflectivity of a single crystal of $YBa_2Cu_3O_7$ and fitted Mattis-Bardeen form of $\delta(\omega)$ (dotted line) in the superconducting state. The arrow shows the peak occurring at 2A ~ 480 cm^{-1}, hence 2A ≈ $8kT_c$, with T_c = 92 K [2.25]. Courtesy of Z. Schlesinger et al.

existence of a gap [2.27]. They also appear to be important for the new class of superconductors. Zero-field nuclear spin lattice relaxation measurements of ^{139}La in $La_{1.8}Sr_{0.2}CuO_{4-\delta}$ below T_c, behave like 1/T, ∝ exp(-A/kT), see Figure 2.8, with activation energy Δ = 1.1 me\' at low temperatures kT ≪ 2Δ due to a T_c = 38 K. A ratio of $2\Delta/kT_c = 7.1$ was obtained [2.28]. Therefore, strong coupling appears to be also present in the La compound. The value of Δ probably has to be attributed to the gap parallel to the planes. In fact, it could be shown that infrared reflectivity data on powders measure the gap along the c-axis, and a ratio of $2\Delta_c/kT_c \approx 2.5$ was given [2.29]. Thus the coupling between the planes would be weak. Such a substantial anisotropic property was not previously found in other superconductors.

From the first measurements of resistivity as a function of magnetic field, the slopes dH_{c2}/dT near T_c, could be obtained, and from them very high critical fields at low temperatures were extrapolated. From the many works published, we quote that of Decroux et *al. [2.30],* also because this was the first paper to

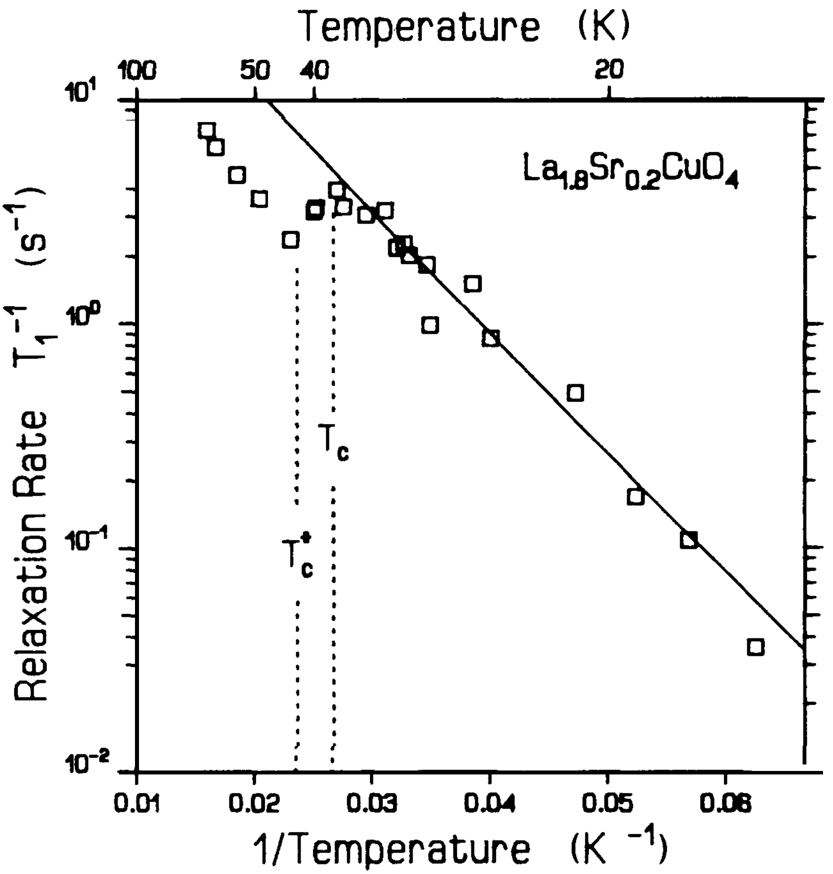


Figure 2.8. Semilogarithmic plot of $1/T_1$ vs 1/T. The straight line demonstrates the activated behavior $1/T_1 \sim \exp[-\Delta/kT]$ for $T \ll T_c$. An activation energy of Δ/k = 135 K is obtained from this graph. From [2.28], ©Les Editions de Physique 1988.

report a specific-heat plateau at T_c. The group at the University of Geneva found $dH_{c2}/dT = -2.5$ T/K, yielding an extrapolated $H_{c2}(T = 0) = 64$ T.

From the well-known formula the critical field in type-II superconductors,

$$H_{c2} = \frac{\phi_0}{2\pi\xi^2}, \tag{2.5}$$

one calculates that the coherence length ξ is of the order of the lattice distances. Actually the coherence lengths evaluated have become smaller; recent results on single crystals by IBM's group in Yorktown Heights [2.31] and the Stanford group on expitaxial layers [2.32] are of the order $\xi_c \approx 3$ - 4 Ω for the coherence length parallel to c and $\xi_{ab} \approx 20$ - 30Ω perpendicular to c.

Such short coherence lengths could be expected when one considers the relation of ξ with the gap and the Fermi energy E_F. Weisskopf [2.33] deduced

$$\xi \approx \frac{E_F}{\Delta} d \tag{2.6}$$

from the Heisenberg uncertainty principle. In Eq. (2.6), d is the screening length, which one can assume to be of the order of a unit-cell distance. The ratio E_F/Δ is near unity owing to the large Δ and the small E_F, the latter resulting from the low carrier density and the sizeable electron mass. Therefore in oxides, ξ is considerably smaller than in metals. Because Δ is anisotropic, so is ξ. The comparable size of E_F and Δ indicates that most of the carriers participate in the superconductivity of the new oxides for temperatures $T < T_c$, in contrast to the classic superconductors, where $E_F \gg \Delta \approx 1.7\ kT_c$.

The short coherence lengths in the layered copper-oxide superconductors are important theoretically, experimentally and applicationwise: The short ξ's and carrier concentrations of the order of $n = 10^{21}/cm^3$ make one wonder whether boson-condensation approaches arc not more appropriate, i.e., real-space Cooper pairing in contrast to the wave-vector space pairing of classical BCS theory [1.10], which applies so well for metals with large ξ's and concentrations n. Actually, Schafroth [2.34] back in 1955 was the first to work out a superconductivity theory with boson condensation. Referring to Chakraverti's phase diagram in Figure 1.2 [1.11], one may regard the metal superconducting phase line as BCS with weak coupling, and the superconducting insulator boundary for large coupling constants λ as the Schafroth line.

The short coherence lengths induce considerable weakening of the pair potential at surfaces and interfaces, as emphasized by Deutscher and Müller [2.24]. Using an expression for the "extrapolation length" b [2.35] for the boundary condition at the superconducting-insulator interface, the $\Delta(x)$ profile was deduced as shown in Figure 2.9 for 1' $\lesssim T_c$ and $T \ll T_c$.

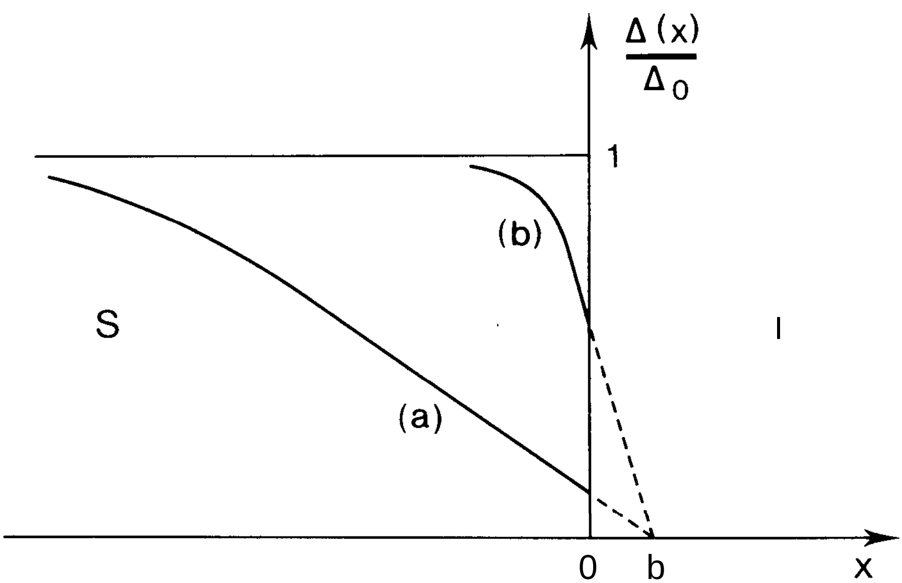


Figure 2.9. Profile of the pair potential in a short-coherence-length superconductor near a superconductor-insulator boundary. Curve a: $T \lesssim T_c$; curve b: $T \ll T_c$. From [2.24], © 1987 The American Physical Society.

Analogous behavior of $\Delta(x)$ will also be present at superconducting-normal (SN) interfaces. Thus, the depressed order parameter involving experiments of SIS and SNS will result in tunneling characteristics [2.24] with a reduced value of the A observed. In consequence, such experiments are less suitable than infrared and NMR to determine Δ, and actually lead to erroneous conclusions regarding gapless superconductivity also in point-contact spectroscopy [2.36]. $YBa_2Cu_3O_7$ undergoes a tetragonal-to-orthorhombic phase transition near 700 C. Thus upon cooling, (110) twin boundaries are formed, separating the orthorhombic domains and inducing ***intragrain*** Josephson or weak-link junctions. These junctions form a network dividing the crystallites into Josephson-coupled domains, with possibility of fluxon trapping as well. Therefore even single crystals can form a superconducting glass in the presence of a sizeable magnetic field.

The basic Hamiltonian regarding the phases is [1.22]

$$\beta H = -\sum J_{ij} \cos(\phi_i - \phi_j - A_{ij}). \tag{2.7}$$

Here J_{ij} is the Josephson coupling constant between domains. The phase factors $A_{ij} = K_{ij}H$ introduce randomness for $H \neq 0$ because K_{ij} is a random geometric factor. A review of the superconducting glass state has recently appeared [2.37].

The first experimental evidence indicating the presence of superconducting glassy behavior was deduced from field-cooled and zero-field-cooled magnetization data [1.24, 2.38] in $La_{2-x}Ba_xCuO_4$ ceramics. In addition to the twin-boundary-induced ***intragrain*** junctions, such a material also has junctions resulting from the ***intragrain*** boundaries. The latter J_{ij}'s are ***much weaker*** and uncouple at lower magnetic fields and currents J_c. Consequently, the critical currents observed in the ceramics are more of the order of 10^3 to 10^4 A/cm^2 [2.39], whereas those in epitaxial layers [2.40] and single crystals [2.1] are of the order of 10^6 to 10^7 A/cm^2 [2.41]. The latter work, carried out by two IBM groups, is a major breakthrough in the field.

The decay length of the superconducting wave functions at SNS and SIS junctions are both of the order of $\xi(0)$. This entails an anomalous temperature dependence of $J_c \propto (T-T_c)^2$. Such behavior is seen in the mid-temperature range for $J_c(T)$ in the $YBa_2Cu_3O_{7-\delta}$ epilayer on $SrTiO_3$ of Figure 2.10 [2.40]. Such critical currents are acceptable for thin-film applications at 77 K for low magnetic fields (Figure 2.10), whereas in the ceramics much lower J_c's require substantial inventiveness or, perhaps better still, a new type of high-T_c superconductor that should exist.

The geometrical critical magnetic field H^*_{c1} is of the order of [1.22]

$$H^*_{c1} = \phi_0/2S, \tag{2.8}$$

where S is the projected area of the superconducting loop with uniform phase. In single crystals, the S of domains is of the order of S = 100 µm^2, whereas that of grains in ceramics is S = 1 - 10 µm^2. In agreement with Eq. (2.8), H^*_{c1} is of

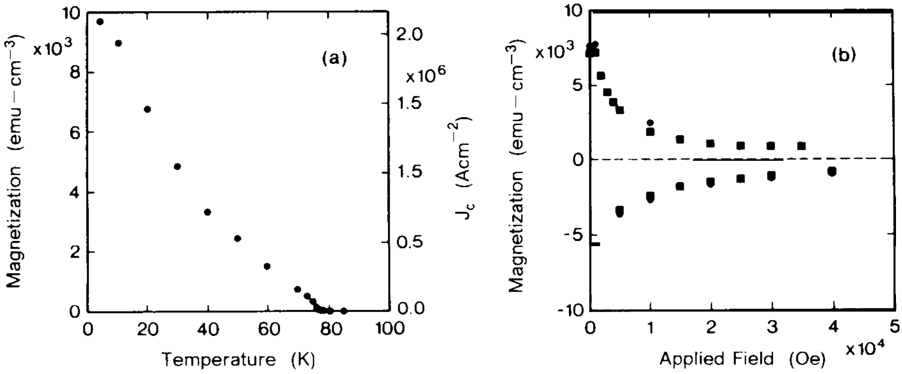


Figure **2.10.** (a) The volume magnetization vs increasing temperature for an epitaxial sample. The y-axis scale on the right was obtained with the use of the Bean formula $J_c = 30$ M/d and with the mean radius of the sample of d = 0.14 cm. (b) The volume magnetization vs applied field at 4.2 K for two samples. From [2.40], © 1987 The American Physical Society.

the order of 0.5 Gauss for the penetration of H into twinned crystals, and 5 to 100 Gauss to disrupt intergranular nets in ceramics.

Since the publication on the existence of this new class of materials, the interest and work have far exceeded the expectations of the laureates, whose aim was primarily to show that oxides could “do better” in superconductivity than metals and alloys. Due to this frenzy, progress on the experimental side has been rapid and is expected to continue. This will also assist in finding new compounds, with T_c‘s reaching at least 130 K (Figure 2.4). Quantitative theoretical models are expected in the not too distant future, first perhaps phenomenological ones. On this rapidly growing tree of research, separate branches are becoming strong, such as glassy aspects, growth techniques for single crystals, epitaxial films, and preparation of ceramics, the latter two being of crucial importance for applications. The former will dominate the small-current microelectronics field, while the latter will have to be mastered in the large-current field. Here the hopes are for energy transport, and large magnetic-field applications for example in beam bending in accelerators and plasma containment in fusion.

References Part 2

[2.1] Cooper, L. N. (1956) Phys. Rev. 104, 1189.

[2.2] Estève, D., Martinis, J. M., Urbina, C., Devoret, M. H., Collin, G., Monod, P., Ribault, M., and Revcolevschi, A. (1987) Europhys. Lett. 3, 1237.

[2.31 Shapiro, S. (1963) Phys. Rev. Lett, *11, 80.*

[2.4] Josephson, B. D. (1962) Phys. Lett, 1, 251.

[2.5] London, F. *(1950) Superfluids,* Vol. I. (John Wiley and Sons, Inc., New York), p. 152.

[2.6] Gough, C. E., Colclough, M. S., Forgan, E. M., Jordan, R. G., Keene, M., Muirhead, C. M., Rae, A. I. M., Thomas, N., Abell, J. S., and Sutton, S. (1987) Nature (London) 326, 855.

[2.7] Steglich, F., Bredl, C. D., De Boer, F. R., Lang hf., Rauchschwalbe, U., Rietschel, H., Schefzyk, R., Sparn, G., and Stewart, G. R. (1987) Physica Scripta *T19, 253.*

[2.8] Fujimori, A., Takayama-Muromachi, E., Uchida, Y., and Okai, B. (1987) Phys. Rev. B 3.5, 8814.

[2.9] Bianconi, A., Castellano, A. C., De Santis, M., Delogu, P., Gargano, A., and Giorgi, R. (1987) Solid State Commun. 63, 1135; Bianconi, A., Castellano, A. C., De Santis, M., Politis, C., Marcelli, A., Mobilio, S., and Savoia, A. (1987) Z. Phys. B 67, 307.

[2.10] Thiry, P., Rossi, G., Petroff, Y., Revcolevschi, A., and Jegoudez, J. (1988) Europhys. Lett, 5, 55.

[2.11] Nücker, M., Fink. J., Fuggle, J. C., Durham, P. J., and Temmerman, W. M. (1988) Phys. Rev. B 37, 5158.

[2.12] Steiner, P., Hüfner, S., Kinsinger, V., Sander, I., Sirgwart, B., Schmitt, H., Schulz, R., Junk, S., Schwitzgebel, G., Gold, A., Politis, C., Müller, H. P., Hoppe, R., Kemmler-Sack, S., and Kunz, C. (1988) Z. Phys. B 69, 449.

[2.13] Shafer, M. W., Penney, T., *and Olson,* B. L. (I 987) Phys. Rev. B 36, 4047.

[2.14] Tarascon, J. M., McKinnon, W. R., Greene, L. H., Hull, G. W., and Vogel, E. M. (1987) Phys. Rev. B 36, 226; see also Bagley, B. G., Greene, L. H., Tarascon, J. M., and Hull, G. W. (1987) Appl. Phys. Lett. *51, 622.*

[2.15] Greene, R. L., Maletta, H., Plaskett, T. S., Bednorz, J. G., and Müller, K. A. (1987) Solid State Commun, 63, 379.

[2.16] Vaknin, D., Sinha, S. K., Moncton, D. E., Johnston, D. C., Newsam, J, M., Safinya, C. R., and King, H. E., Jr. (1987) Phys. Rev. Lett. 58, 2802.

[2. 17] Shirane, G., Endoh, Y., Birgeneau, R. J., Kastner, M. A., Hidaka, Y., Oda, M., Suzuki, M., and Murakami, T. (1987) Phys. Rev. Lett. 59, 1613.

[2.18] Emery, V. J. (1987) Phys. Rev. Lett. 58, 2794; *idem* (1987) Nature (London) 328, 756.

[2.19] Hirsch, J. E. (1987) Phys. Rev. Lett. 59, 228; Cyrot, M. (1987) Solid State Commun. 62, 821, and to be published.

[2.20] Anderson, P. W. (1987) Science 235, 1196; Anderson, P. W., Baskaran, G., Zou, Z., and Hsu. T. (1987) Phys. Rev. Lett. 58, 2790.

[2.21] Batlogg, B., Cava, R. J., Jayaraman, A., Van Dover, R. B., Kourouklis, G. A., Sunshine, S., Murphy, D. W., Rupp, L. W., Chen, H. S., White, A., Short, K. T., Mujsce, A. M., and Rietman, E. A. (1987) Phys. Rev. Lett. 58, 2333; Bourne, L. C., Crommite, M. F., Zettl, A., zur Loye, H.-C., Keller, S. W., Leary, K. L., Stacy, A. M., Chang, K. J., Cohen, M. L., and Morris, D. E. (1987) *ibid, 2337.*

[2.22] Faltens, T. A., Ham, W. K., Keller, S. W., Leary, K. J., Michaels, J. N., Stacy. A. M., zur Loye, H.-C., Morris, D. E., Barbee III, T. W., Bourne, L. C., Cohen, M. L., Hoen, S., and Zettl, A. (1987) Phys. Rev. Lett. 59,915; Batlogg , B., Kourouklis, G., Weber, W., Cava, R. J., Jayaraman, A., White, A. E., Short, K. T., Rupp, L. W., and Rietman, E. A. (1987) *ibid,* 912.

[2.23] Leary, K. J., zur Loye, H.-C., Keller, S. W., Faltens, T. A., Ham, W. K., Michaels, J. N., and Stacy, A. M. (1987) Phys. Rev. Lett. 59, 1236.

[2.24] Deutscher, G. and Müller, K. A. (1987) Phys. Rev. Lett. 59, 1745.

[2.25] Schlesinger, Z., Collins, R. T., Kaiser, D. L., and Holtzberg, F., (1987) Phys. Rev. Lett. 59, 1958.

[2.26] Mali, M., Brinkmann, D., Pauli, L., Roos, J., Zimmermann, H., and Hulliger, J. (1987) Phys. Lett. A 124, 112, and references therein. It should be noted that analysis of the relaxation data $1/T_c \propto T^3$ is also possible, as one expects in the presence of anisotropic gaps.

[2.27] Hebel, L. C. and Slichter, C. P. (1959) Phys. Rev. *113*, 1504.

[2.28] Seidel, H., Hentsch, F., Mehring, M., Bednorz, J. G., and Müller, K. A. (1988) Europhys. Lett. 5, 647.

[2.29] Schlesinger, Z., Collins, R. T., and Shafer, M. W‘. (1987) Phys. Rev. B 35, 7232.

[2.30] Decroux, M., Junod, A., Bezinge, A., Cattani, D., Cors, J., Jorda, J. L. Stettler, A., Francois, M., Yvon, K., Fischer, Ø., and Muller, J. (1987) Europhys. Lett. 3, 1035.

[2.31] Worthington, T. K., Gallagher, W. J., and Dinger, T. R. (1987) Phys. Rev. Lett. 59, 1160.

[2.32] Kapitulnic, A., Beasley, M. R., Castellani, C., and Di Castro, C. (1988) Phys. Rev. B 37, 537.

[2.33] Weisskopf, V. F. (1979) “The Formation of Cooper Pairs and the Nature of Superconducting Currents,” CERN, Geneva, Memorandum No. CERN 79-12, Theoretical Studies Division, Dec. 2 1, 1979.

[2.34] Schafroth, M. R., (1954) Phys. Rev. 96, 1149 and 1442; *idem* (1955) Phys. Rev. *100, 463.*

[2.35] de Gennes, P. *G.* (1968) *Superconductivity of Metals and Alloys* (Benjamin, New York) p. 229.

[2.36] Yanson, I. K., Rybal’chenko, L. F., Fisun, V. V., Bobrov, N. L., Obolenskii, M. A., Brandt, N. B., Moshchalkov, V. V., Tret’yakov, Yu. D., Kaul’, A. R., and Graboi, I. É. (1987) Fiz. Nizk. Temp. 13,557 [Sov. J. Low Temp. Phys. 13, 315 (1987)]

[2.37] Müller, K. A., Blazey, K. W., Bednorz, J. G., and Takashige, M. (1987) Physica 148 B, 149.

[2.38] Razavi, F. S., Koffyberg, F. P., and Mitrović, B. (1987) Phys. Rev. B 35, 5323.

[2.39] Kwak, J, F., Venturini, E. L., Ginley, D. S., and Fu, W. (1987) Novel *Superconductiviy, Proceedings of the International Workshop on Novel Mechanisms in Superconductivity,* editors, S. A. Wolf and V. Z. Kresin (Plenum Press, New York) p. 983.

[2.40] Chaudhari, P., Koch, R. H., Laibowitz, R. B., McGuire, T. R., and Gambino, R. J, (1987) Phys. Rev. Lett. 58, 2684.

[2.41] Dinger, T. R., Worthington, T. K., Gallagher, W. J., and Sandstrom, R. L. (1987) Phys. Rev. Lett. 58, 2687.

Physics 1988

LEON M LEDERMAN, MELVIN SCHWARTZ and JACK STEINBERGER

for the neutrino beam method and the demonstration of the doublet structure of the leptons through the discovery of the muon neutrino

THE NOBEL PRIZE IN PHYSICS

Speech by Professor Gösta Ekspong of the Royal Swedish Academy of Sciences.
Translation from the Swedish text.

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen,

The Royal Swedish Academy of Sciences has decided to award this year's Nobel Prize in Physics jointly to Dr Leon Lederman, Dr Melvin Schwartz and Dr Jack Steinberger. The citation has the following wording, "for the neutrino beam method and the demonstration of the doublet structure of the leptons through the discovery of the muon neutrino".

The neutrino figures in George Gamow's entertaining book "Mr Tompkins Explores the Atom", written in the 1940's. Gamow describes how Mr Tompkins in a dream visits a woodcarvers shop, where the building blocks of the elements-protons, neutrons and electrons- are stored in separate caskets. Mr Tompkins sees many unusual things, but above all a carefully closed, but apparently empty casket labelled: "NEUTRINOS, Handle with care and don't let out". The woodcarver does not know whether there is anything inside. The friend, who had presented the casket to him, must have been Wolfgang Pauli, Nobel Laureate in Physics in 1945, who proposed the existence of the neutrino in the early 1930's.

The neutrino is electrically neutral and almost or totally massless - hence the name. It cannot be seen and it interacts only weakly with atoms. It travels with the speed of light or nearly so. It is impossible to completely stop a beam of neutrinos. To do so would require a wall of several hundred thousands of steel blocks stacked in depth one after the other, each with a thickness corresponding to the distance from here to the sun.

Our sun is a source of neutrinos, which are copiously produced in its hot central region. They pass through the whole sun without much difficulty. Every square centimeter on Earth is bombarded by many billion solar neutrinos every second and they pass straight through the Earth without leaving a noticable mark. The neutrinos are-if I may say so- "lazy", they do almost nothing but steal energy, which they carry away.

The great achievement of the Nobel prize winners was to put the "lazy" neutrinos to work. Lederman, Schwartz and Steinberger are famous for several other important discoveries concerning elementary particles. At the time of the neutrino experiment they were associated with Columbia University in New York. They and their co-workers designed the world's first beam of neutrinos at the Brookhaven National Laboratory, using its large proton accelerator as a source. Their neutrinos had considerably more energy than usual, because they were produced from the decay in flight of fast moving mesons. Such neutrinos are much more apt to interact with matter and the collisions with atomic particles become much more interesting. Although a neutrino collision is a rare event it can be spectacular at high energy- and very informative.

In their pioneering experiment the prizewinners dealt with a total of about 10^{14}, i.e. a hundred thousand billion neutrinos. To catch just a few dozen collisions from all these, the research team invented and built a huge, sophisticated detector with the weight of 10 tons. Other unwanted particles in the beam had to be prevented from entering the detector. An enormous 13-meter thick steel wall served this purpose. To save time and money, the wall material was taken from scrapped battleships. Unwanted particles came also from the outside in the form of cosmic ray muons. Various tricks were used to prevent these muons from playing a role as false neutrinos. The first neutrino beam experiment was a bold endeavor, which proved successful. The method has since been much used as a tool for investigating the weak force and the quark structure of matter. It has also been used to investigate the neutrino itself.

At the time of the prizewinners' experiment, physicists were puzzled by the fact that a possible, alternative decay of the muon particle did not happen. No known law forbade it, and there is a general principle which says that a process must occur unless it is explicitly forbidden by law. The mystery was solved when the prizewinners' team discovered that Mother Nature provides two completely different species of neutrino, as had been suggested by a theoretical analysis. The old type of neutrino is paired with and may be transformed into an electron, the newly discovered type of neutrino is similarly paired with the muon. The two pairs constitute two separate lepton families, which never mix with each other. Thus, a new law of Nature had been discovered.

Cosmologists and physicists alike want to know how many different lepton families, i.e. how many neutrino species there are in Nature. Present ideas about the birth and early evolution of our universe cannot tolerate more than four. A third is already on the books. One of the goals of the experimental program at the large LEP accelerator ring at CERN, which will be ready to start operation next summer, is to give a precise answer as to the number of neutrino species and thus the exact number of lepton families in the universe.

Professors Dr Lederman, Dr Schwartz and Dr Steinberger,
You started a bold new line of research, which gave rich fruit from the beginning by establishing the existence of a second neutrino. Furthermore, problems which could not even be formulated at the time of your experiment, have been successfully elucidated in later experiments using your method. The pairing of the leptons, which you discovered, is also of much wider applicability than could be foreseen at the time and is now an indispensible ingredient in the standard model for quarks and leptons.

On behalf of the Royal Swedish Academy of Sciences, I have the privilege and the great honour to extend to you our warmest congratulations. May I now ask you to receive the 1988 Nobel Prize in physics from the hands of His Majesty the Ring.

MELVIN SCHWARTZ

Having been born in 1932, at the peak of the great depression, I grew up in difficult times. My parents worked extraordinarily hard to give us economic stability but at the same time they managed to instill in me two qualities which became the foundation of my personal and professional life. One is an unbounded sense of optimism; the other is a strong feeling as to the importance of using one's mind for the betterment of mankind.

My interest in Physics really began at the age of 12 when I entered the Bronx High School of Science in New York. That school has become famous for the large number of outstanding individuals it has produced including among them four Nobel Laureates in Physics. The four years I spent there were certainly among the most exciting and stimulating of my life, mostly because of the interaction with other students having similar background, interest and ability. It's rather amazing how important the interaction with the one's peers can be at that age in determining one's direction and success in life.

Upon graduating from high school the path to follow was fairly obvious. The Columbia Physics Department at that time was unmatched by any in the world. Largely a product of the late Professor I.I. Rabi, it was a department which was to provide the ambiance for six Nobel Prize pieces of work in widely diverse fields during the next thirteen years. And, in addition, it was the host for a period of time to another half dozen or so future Nobel Laureates either as students or as post-doctoral researchers. I know of no other institution either before or since that has come close to that record.

Thus, it was that I became an undergraduate at Columbia in 1949, to stay there through my graduate years and take up a faculty position as Assistant Professor in 1958. I became an Associate Professor in 1960 and a Professor in 1963.

In order for me to put my life into perspective, I must mention four individuals who have given it meaning, 'direction and focus. Foremost among these is my wife Marilyn whom I married 35 years ago and who has provided the one most enduring thread throughout these years. Without her constant encouragement and enthusiasm there would have been far less meaning to my life. The second is of course Jack Steinberger. Jack was my teacher, my mentor and my closest colleague during my years at Columbia. Whatever taste and judgement I have ever had in the field of Particle Physics came from Jack. Third of course is T.D. Lee. He was the inspirer of this experiment and the person who has served as a constant sounding board for any ideas I have had. He has also become, I am proud to say, a dear personal friend. And finally, my close collaborator Leon Lederman. If there is any one person who has served as the sparkplug for high energy physics in the U.S. it has been Leon. I am proud to have been his collaborator.

In 1966, after having spent 17 years at Columbia, I decided to move West to Stanford, where a new accelerator was just being completed. During the ensuing years I was involved in two major research efforts. The first of these investigated the charge asymmetry in the decay of the long-lived neutral kaon. The second of these, which was quite unique, succeeded in producing and detecting relativistic hydrogen-like atoms each made up of a pion and a muon.

During the 1970's, lured in part by the new industrial revolution in "Silicon Valley" I decided to try my hand at a totally new adventure. Digital Pathways, Inc. of which I am currently the Chief Executive Officer is a company dedicated to the secure management of data communications. Although it is difficult to predict the future I still have all the optimism that I had back when I first grew up in New York - life can be a marvelous adventure.

(added in 1991): A new change in my career occurred in February 1991 when I became Associate Director, High Energy and Nuclear Physics, at Brookhaven National Laboratory.

THE FIRST HIGH ENERGY NEUTRINO EXPERIMENT

Nobel Lecture, December 8, 1988

bY

MELVIN SCHWARTZ

Digital Pathways, Inc., 201 Ravendale Avenue, Mountain View, CA 94303, USA.

In the first part of my lecture I would like to tell you a bit about the state of knowledge in the field of Elementary Particle Physics as the decade of the 1960's began with particular emphasis upon the Weak Interactions. In the second part I will cover the planning, implementation and analysis of the first high energy neutrino experiment. My colleagues, Jack Steinberger and Leon Lederman, will discuss the evolution of the field of high energy neutrino physics beyond this first experiment and the significance of this effort when seen in the context of today's view of elementary particle structure.

I. HISTORICAL REVIEW

By the year 1960 the interaction of elementary particles had been classified into four basic strengths. The weakest of these, the gravitational interaction does not play a significant role in the laboratory study of elementary particles and will be ignored. The others are:

1. *Strong Interactions*

This class covers the interactions among so-called hadrons. Among these hadrons are the neutrons and protons that we are all familiar with along with the pions and other mesons that serve to tie them together into nuclei. Obviously, the interaction that ties two protons into a nucleus must overcome the electrostatic repulsion which tends to push them apart. The strong interactions are short range, typically acting over a distance of 10^{-13} cm, but at that distance are some two orders of magnitude stronger than electromagnetic interactions.

In general, as presently understood, hadrons are combinations of the most elementary strongly interacting particles, called quarks. You will hear more about them later.

2. *Electromagnetic Interactions*

You are all familiar with electromagnetic interactions from your daily experience. Like charges repel one another. Opposite charges attract. The earth acts like a giant magnet. Indeed matter itself is held together by the electromagnetic interactions among electrons and nuclei. With the exception of the neutrinos, all elementary particles have electromagnetic interactions either through charge, or magnetic property, or the ability to directly interact with charge or magnetic moment. In 1960, the only known elementary particles apart from the hadrons were the three leptons — electron, muon and neutrino with some suspicion that there might be two types of neutrinos. Both the electron and muon are electromagnetically interacting.

3. Weak *Interactions*

Early in the century it was discovered that some nuclei are unstable against decay into residual nuclei and electrons or positrons. There were two important characteristics of these so-called decays.

a. They were "slow". That is to say, the lifetimes of the decaying nuclei corresponded to an interaction which is much weaker than that characteristic of electromagnetism.

b. Energy and momentum were missing.

If one examined the spectrum of the electrons which were emitted, then it was clear that to preserve energy, momentum and angular momentum in the decay it was necessary that there be another decay product present. That decay product needed to be of nearly zero mass and have half integral spin. This observation was first made by Pauli. Fermi later gave it the name of neutrino.

The development of the Fermi theory of weak interactions in fact made the neutrino's properties even more specific. The neutrino has a spin of 1/2 and a very low probability of interacting in matter. The predicted cross-section for the interaction of a decay neutrino with nucleons is about 10^{-43} cm^2. Thus, one of these neutrinos would on the average pass through a light year of lead without doing anything.

The B-decay reactions can be simply written as:

$$Z \rightarrow (Z+1) + e^- + \bar{\nu}$$
$$Z \rightarrow (Z,-1) + e+ + v$$

By the failure to detect neutrino-less double β decay, namely the process $Z \rightarrow$ (Z-l-2) + $e^- + e^-$, it was established that the neutrino and anti-neutrino were indeed different particles. In the 1950's, by means of a series of experiments associated with the discovery of parity violation it was also established that the neutrinos and anti-neutrinos were produced in a state of complete longitudinal polarization or helicity, with the neutrinos being left-handed and anti-neutrinos right-handed.

In the 1940's and 1950's, a number of other weak interactions were discovered. The pion, mentioned earlier as the hadron which serves to hold the nucleus together, can be produced in a free state. Its mass is about 273 times the electron mass and it decays in about 2.5 X 10^{-8} seconds into a muon and a particle with neutrino-like properties. The muon in turn exhibits all of the properties of a heavy electron with a mass of about 207 times the electron mass. It decays in about 2.2 X 10^{-6} seconds into an electron and two neutrinos. The presumed reactions, when they were discovered, were written as:

$$\pi^+ \rightarrow \mu^+ + \nu$$
$$\pi^- \rightarrow \mu^- + \bar{\nu}$$
$$\mu^+ \rightarrow e^+ + \nu + \bar{\nu}$$

It was also known by 1960 that these decays were parity violating and that the neutrinos here had the same helicity as the neutrinos emitted in β decay.

Needless to say, there was a general acceptance in 1959 that the neutrinos associated with β decay were the same particles as those associated with pion and muon decay. The only hint that this may not be so came from a paper by G. Feinberg in 1958 in which he showed that the decay $\mu \rightarrow e + \gamma$ should occur with a branching ratio of about 10^{-4} if a charged intermediate boson (W) moderated the weak interaction. Inasmuch as the experimental limit was much lower (~ 10^{-8}) this paper was thought of as a proof that there was no intermediate boson. Feinberg did point out, however, that a boson might still exist if the muon neutrino and the electron neutrino were different.

One final historical note with respect to neutrinos. In the mid-nineteen fifties Cowen and Reines in an extremely difficult pioneering experiment were able to make a direct observation of the interaction of neutrinos in matter. They used a reactor in which a large number of $\bar{\nu}$ are produced and observed the reaction $\bar{\nu} + p \rightarrow n + e^+$. The cross-section observed was consistent with that which was required by the theory.

II. CONCEPTION, PLANNING AND IMPLEMENTATION OF THE EXPERIMENT

The first conception of the experiment was in late 1959. The Columbia University Physics Department had a tradition of a coffee hour at which the latest problems in the world of Physics came under intense discussions. At one of these Professor T. D. Lee was leading such a discussion of the possibilities for investigating weak interactions at high energies. A number of experiments were considered and rejected as not feasible. As the meeting broke up there was some sense of frustration as to what could ever be done to disentangle the high energy weak interactions from the rest of what takes place when energetic particles are allowed to collide with targets. The only ray of hope was the expectation that the cross sections characteristic of the weak interactions increased as the square of the center of mass energy at least until such time as an intermediate boson or other damping mechanism took hold.

That evening the key notion came to me - perhaps the neutrinos from pion decay could be produced in sufficient numbers to allow us to use them in an experiment. A quick "back of the envelope" calculation indicated the feasibility of doing this at one or another of the accelerators under construction or being planned at that time. I called T. D. Lee at home with the news and his enthusiasm was overwhelming. The next day planning for the experiment began in earnest. Meanwhile Lee and Yang began a study of what could be learned from the experiment and what the detailed cross-sections were.

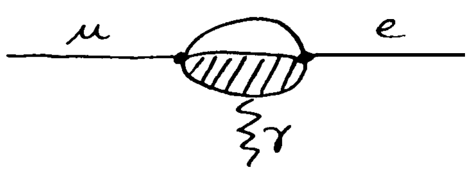
Not long after this point we became aware that Bruno Pontecorvo had also come up with many of the same ideas as we had. He had written up a proposed experiment with neutrinos from stopped pions, but he had also discussed the possibilities of using energetic pions at a conference in the Soviet Union. His overall contribution to the field of neutrino physics was certainly major.

Leon Lederman, Jack Steinberger, Jean-Marc Gaillard and I spent a great deal of time trying to decide on an ideal neutrino detector. Our first choice, if it were feasible, would have been a large Freon bubble chamber that Jack Steinberger had built. (In the end that would have given about a factor of 10 fewer events at the Brookhaven A.G.S. than the spark chamber which we did use. Hence it was not used in this experiment).

Fortunately for us, the spark chamber was invented at just about that time. Gaillard, Lederman and I drove down to Princeton to see one at Cronin's laboratory. It was small, but the idea was clearly the right one. The three of us decided to build the experiment around a ten ton spark chamber design.

In the summer of 1960, Lee and Yang again had a major impact on our thinking. They pointed out that it was essentially impossible to explain the absence of the decay $\mu \rightarrow e + \gamma$ without positing two types of neutrinos. Their argument as presented in the 1960 Rochester Conference was more or less as follows:

1. The simple four-fermion point model which explains low energy weak interactions leads to a cross-section increasing as the square of the center of mass energy.
2. At the same time, a point interaction must of necessity be S-wave and thus the cross-section cannot exceed $\lambda^2/4\pi$ without violating unitarity. This violation would take place at about 300 GeV.
3. Thus, there must be a mechanism which damps the total cross-section before the energy reaches 300 GeV. This mechanism would imply a "size" to the interaction region which would in turn imply charges and currents which would couple to photons. This coupling would lead to the reaction $\mu \rightarrow e + \gamma$ through the diagram.



4. The anticipated branding ratio for $\mu \rightarrow e + \gamma$ should not differ appreciably from 10^{-5}. The fact that the branching ratio was known to be less than 10^{-8} was then *strong evidence* for the two-neutrino hypothesis.

With these observations in mind the experiment became highly motivated toward investigating the question of whether $\nu_\mu = \nu_e$. If there were only one type of neutrino then the theory predicted that there should be equal numbers of muons and electrons produced. If there were two types of neutrinos then the production of electrons and muons should be different. Indeed, if one followed the Lee-Yang argument for the absence of $\mu \rightarrow e + \gamma$ then the muon neutrino should produce *no* electrons at all.

We now come to the design of the experiment. The people involved in the effort were Gordon Danby, Jean-Marc Gaillard, Konstantin Goulianos, Nariman Mistry along with Leon Lederman, Jack Steinberger and myself. The facility used to produce the pions was the newly completed Alternate Gradient Synchrotron (A.G.S.) at the Brookhaven National Laboratory. Although the maximum energy of the accelerator was 30 GeV, it was necessary to run it at 15 GeV in order to minimize the background from energetic muons.

Pions were produced by means of collisions between the internal proton beam and a beryllium target at the end of a 3-meter straight section (see Figure 1). The detector was set at an angle of 7.5° to the proton direction behind a 13.5-meter steel wall made of the deck-plates of a dismantled cruiser. Additional concrete and lead were placed as shown.

To minimize the amount of cosmic ray background it was important to minimize the fraction of time during which the beam was actually hitting the target. Any so-called "events" which occurred outside of that window could then be excluded as not being due to machine induced high energy radiation.

The A.G.S. at 15 GeV operator at a repetition rate of one pulse per I.2 seconds. The beam RF structure consisted of 20 ns bursts every 220 ns. The beam itself was deflected onto the target over the course of 20-30 μs for each cycle of the machine. Thus, the target was actually being bombarded for only 2×10^{-6} sec. for each second of real time.

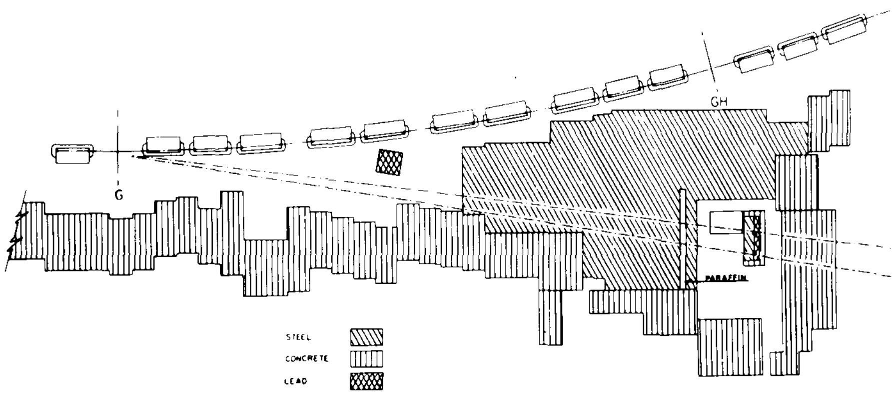


Figure 1. Plan view of the A.G.S. neutrino experiment.

In order to make effective use of this beam structure it was necessary to gate the detector on the bursts of pions which occurred when the target was actually being struck. This was done by means of a 30 ns time window which was triggered through the use of a Cerenkov counter in front of the shielding wall. Phasing of the Cerenkov counter relative to the detector was accomplished by raising the A.G.S. energy and allowing muons to penetrate the shield.

Incidentally, this tight timing also served to exclude 90 % of the background induced by slow neutrons.

The rate of production of pions and kaons was well known at the time and it was quite straightforward to calculate the anticipated neutrino flux. In Figure 2 we present an energy spectrum of the neutrino flux for a 15 GeV proton beam making use of both pion and kaon decay. It is clear that kaon decay is a major contributor for neutrino energies greater than about 1.2 GeV. (These neutrinos come from the reaction $K^{\pm} \rightarrow \mu^{\pm} + (\frac{\nu}{\bar{\nu}})$).

Needless to say, the main shielding wall is thick enough to suppress all strongly interacting particles. Indeed, the only hadrons that were expected to emerge from that wall were due to neutrino interactions in the last meter or so. Muons entering the wall with up to 17 GeV would have been stopped by ionization loss. The only serious background was due to neutrons leaking through the concrete floor; these were effectively eliminated in the second half the experiment.

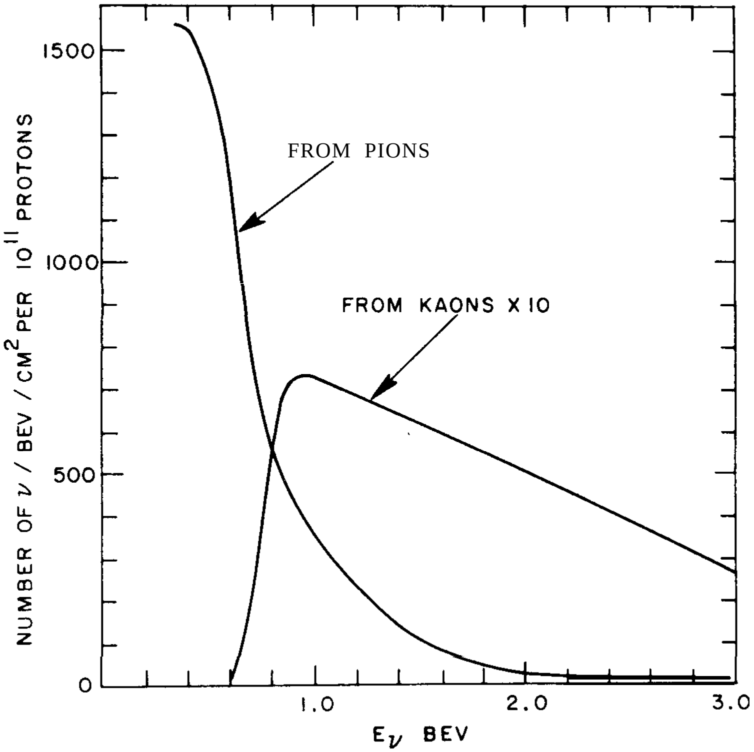


Figure 2. Energy spectrum of neutrinos as expected for A.G.S. running at 15 GeV.

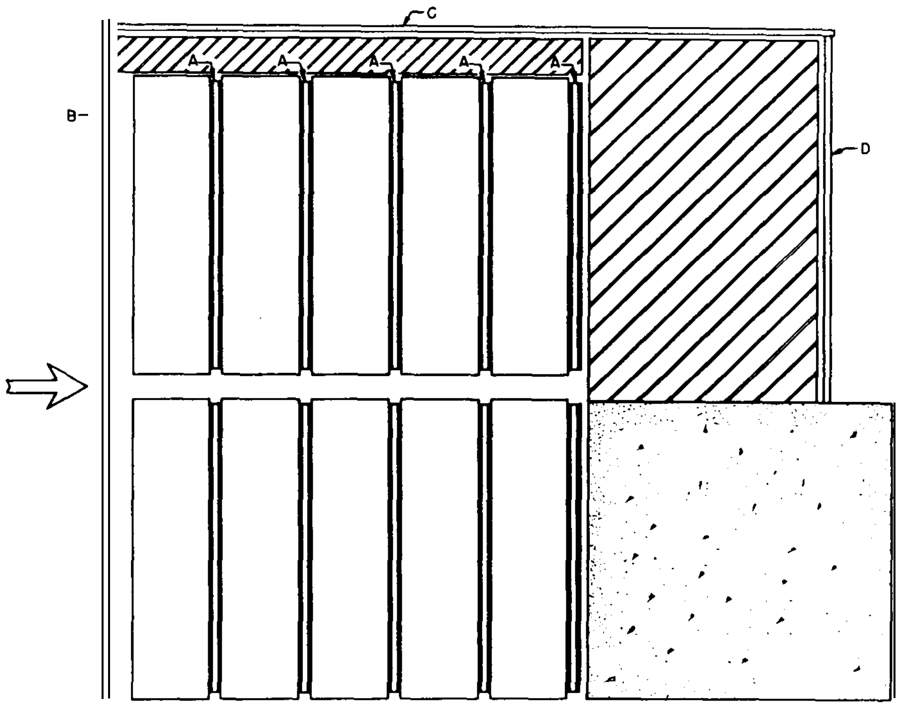


Figure 3. Spark chambcr and countcr arrangement. This is the front view with neutrinos cntcring on the left. A arc the trigger counters. B, C and D are used in anti-coincidcnce.

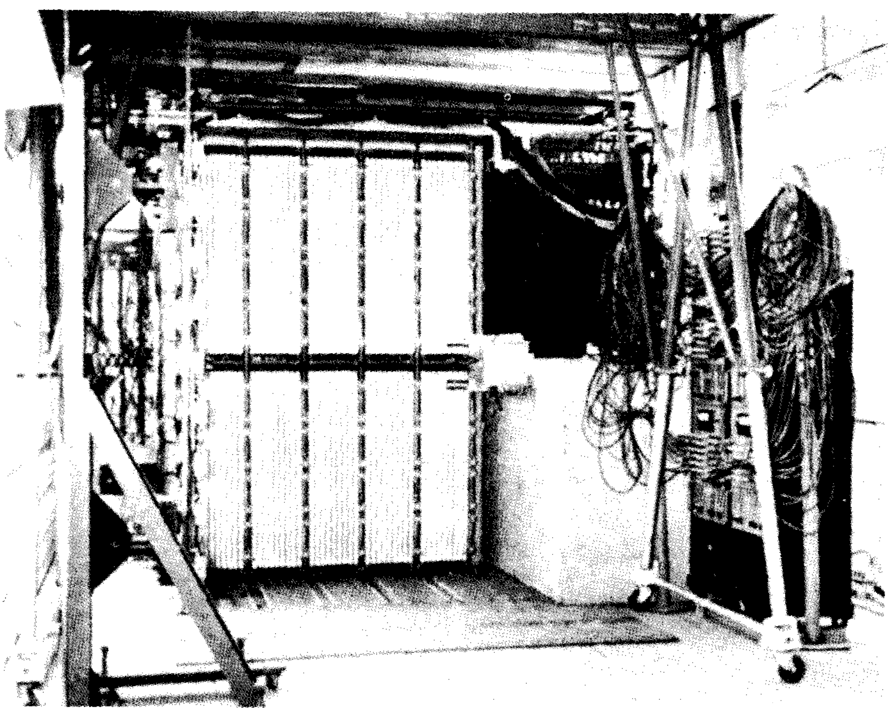
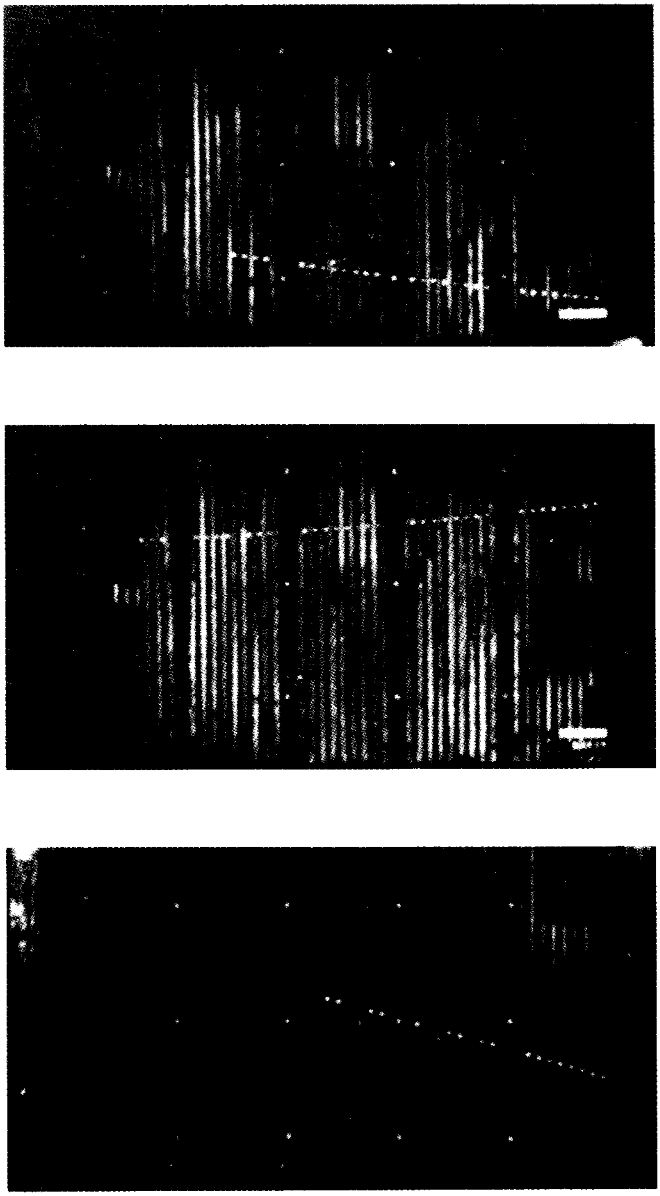


Figure 4. A photograph of the chambers and counters

The spark chamber is shown in Figure 3 and 4. It consisted of ten modules, each of 9 aluminum plates, 44 in. x 44 in. x 1 in. thick separated by 3/8 in. Lucite spacers. Anticoincidence counters covered the front, top and rear of the assembly, as shown, to reduce the effect of cosmic rays and muons which penetrate the shielding wall. Forty triggering counters were inserted between modules and at the end of the assembly. Each triggering counter consisted of two sheets of scintillator separated by 3/4 in. of aluminum. The scintillators were put in electronic coincidence.



A

B

C

Figure 5. Some typical single muon events.

Events were selected for further study if they originated within a fiducial volume which excluded the first two plates, two inches at top and bottom and four inches at front and rear of the assembly. Single track events also needed to stay within the fiducial volume if extrapolated back for two gaps. Single tracks were not accepted unless their production angle relative to the neutrino direction was less than 60”.

A total of 113 events were found which satisfied these criteria. Of these, 49 were very short single tracks. All but three of these appeared in the first half of the experiment before the shielding was improved and they were considered to be background. In retrospect, some of these were presumably neutral current events, but at the time it was impossible to distinguish them from neutron induced interactions due to leakage over and under the shield.

The remaining events included the following categories.

a) 34 “single muons” of more than 300 MeV/c of visible momentum. Some of these are illustrated in Figure 5. Among them are some with one or two extraneous sparks at the vertex, presumably from nuclear recoils.

b) 22 “vertex” events. Some of these show substantial energy release. These events are presumably muons accompanied by pions in the collision. (See Figure 6)

c) 8 “shower” candidates. Of these 6 were selected so that their potential range, had they been muons would correspond to more than the 300 MeV/c. These were the only candidates for single electrons in the experiment. We will consider them in detail shortly.

It was quite simple to demonstrate that the 56 events in categories (a) and (b) were almost all of neutrino origin.

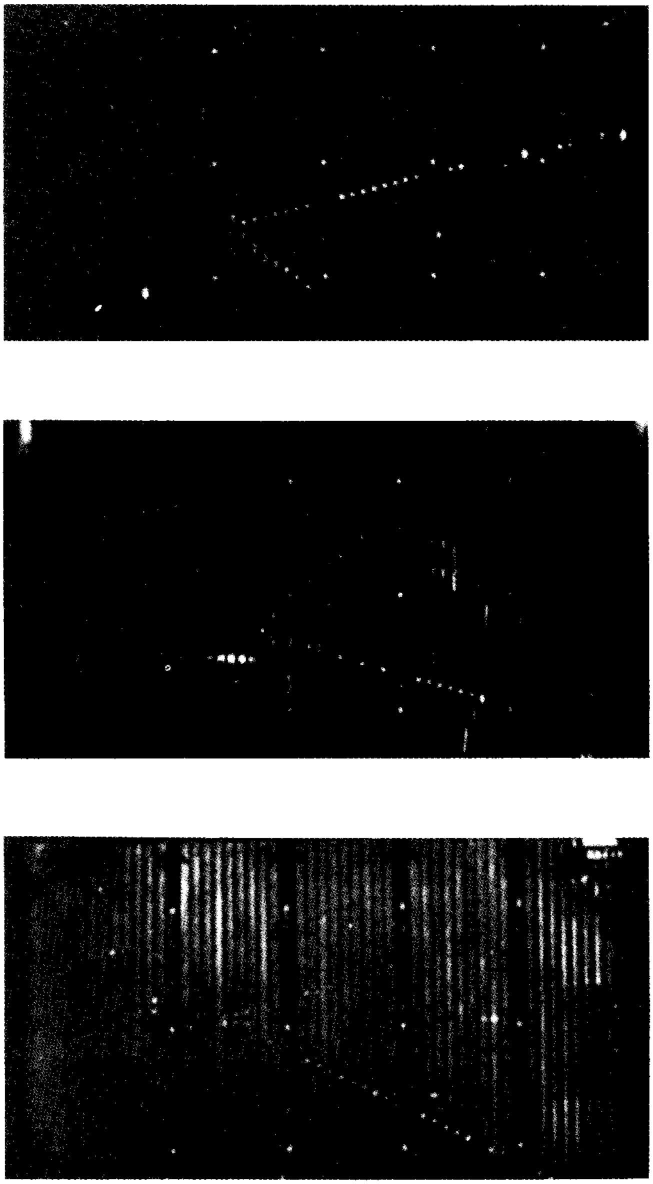
By running the experiment with the accelerator off and triggering on cosmic rays it was possible to place a limit of 5 ± 1 on the total number of the single muon events which could be due to such background. Indeed, the slight asymmetry in Figure 7 is consistent with this hypothesis.

It was simple to demonstrate that these events were not neutron induced. Referring to Figure 7 we see how they tend to point toward the target through the main body of steel shielding. No more than 10^{-4} events should have arisen from neutrons penetrating the shield (other than from neutrino induced events in the last foot of the shield itself). Indeed, removing four feet of steel from the front would have increased the event rate by a factor of 100; no such increase was seen. Futhermore, if the events were neutron induced they would have clustered toward the first chambers. In fact they were uniformly spread throughout the detector subject only to the 300 MeV/c requirement.

The evidence that the single particle tracks were primarily due to muons was based on the absence of interactions. If these tracks were pions we would have expected 8 interactions. Indeed, even if all of the stopping tracks were considered to be interacting, it would still lead to the conclusion that the mean free path of these tracks was 4 times that expected for hadrons.

As a final check on the origin of these events we effectively replaced four feet of the shield by an equivalent amount as close as possible to the beryllium target. This reduced the decay distance by a factor of 8. The rate of events decreased from 1.46 ± .02 to 0.3±0.2 per 10^{16} incident protons.

All of the above arguments convinced us that we were in fact looking at neutrino induced events and that 29 of the 34 single track events were muons produced by neutrinos (The other five being background due to cosmic rays). It is these events that will form the basis of our arguments as to



A

B

C

Figure 6. Some typical “vertex” events.

the identity of ν_μ and ν_e. But, first we must see what electrons would look like in passing through our spark chambers. An electron will on the average radiate half of its energy in about four of the aluminum plates. This will lead to gammas which will in turn convert to other electron -positron pairs. The net result is called a "shower". Typically an electron shower shows a number of sparks in each gap between plates. The total number of sparks in the shower increases roughly linearly with electron energy in 400 MeV region.

In order to calibrate the chambers we exposed them to a beam of 400 GeV electrons at the Brookhaven Cosmotron (See Figure 8). We noted that the triggering system was 67 % efficient with respect to these electrons. We then plotted the spark distribution as shown in Figure 9 for a sample of 2/3 × 29 expected showers. The 6 "shower" events were also plotted. Clearly, the difference between the expected distribution, had there been only one neutrino, and the observed distribution was substantial. We concluded that $\nu_\mu \neq \nu_e$.

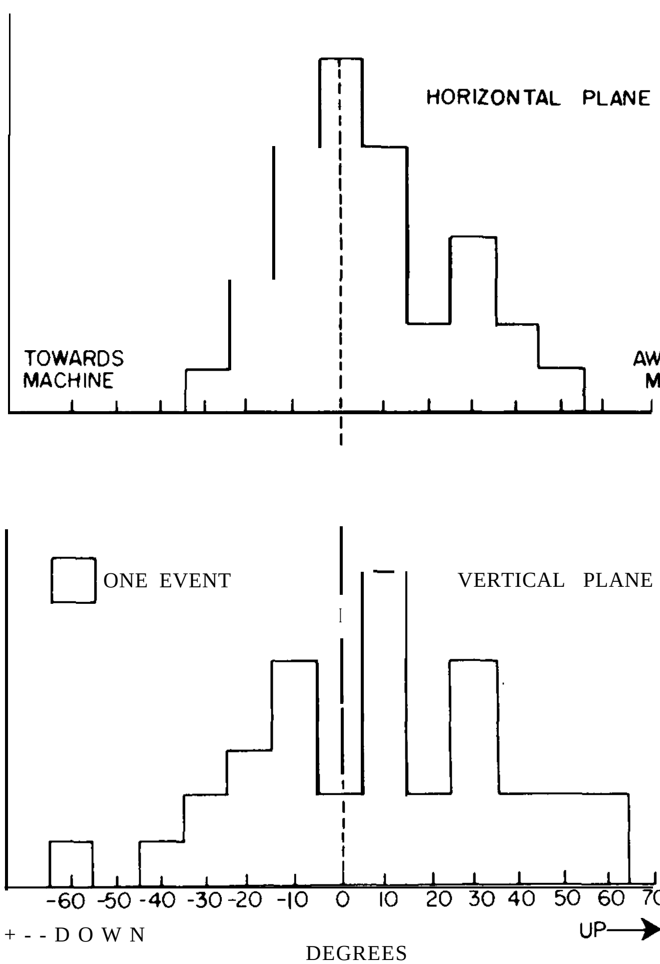


Figure 7. Projected angular distribution of the single track events. The neutrino direction is taken as zero degrees.

As a further point, we compared the expected rate of neutrino events with that predicted by the Fermi theory and found agreement within 30 %.

The results of the experiment were described in an article in Physical Review Letters Volume 9, pp. 36-44 (1962).

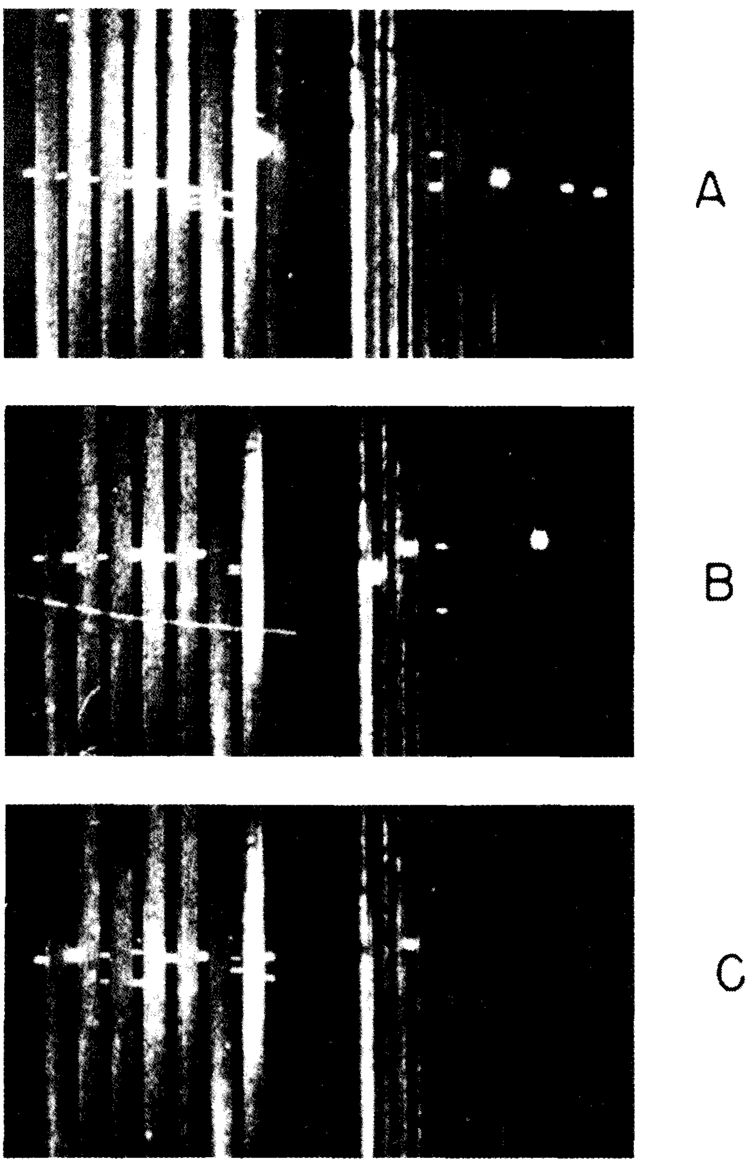


Figure 8. Typical 400 McV/c clectrons from the Cosmotron calibration run.

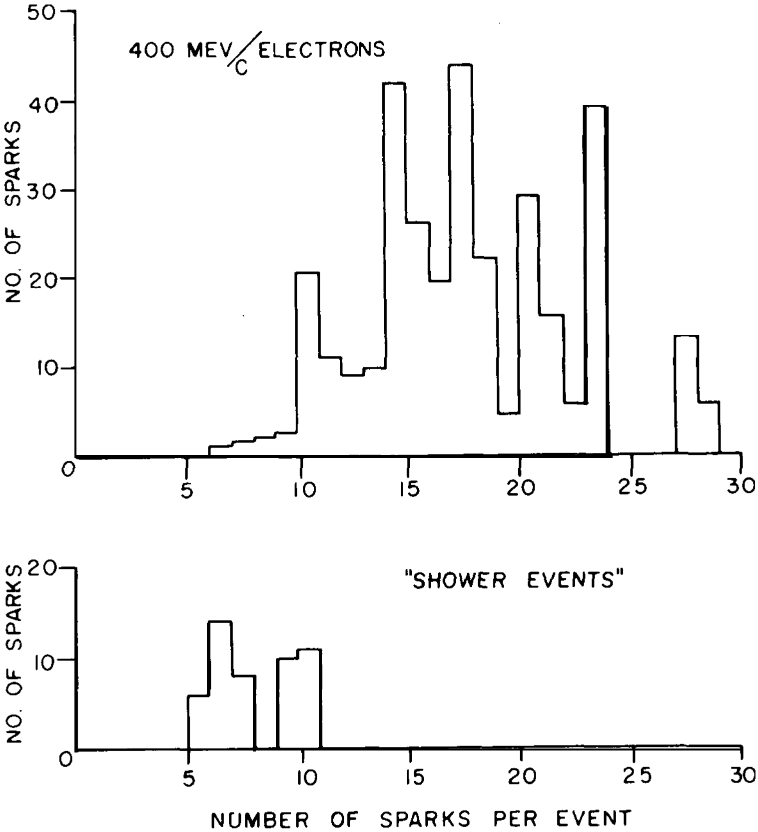
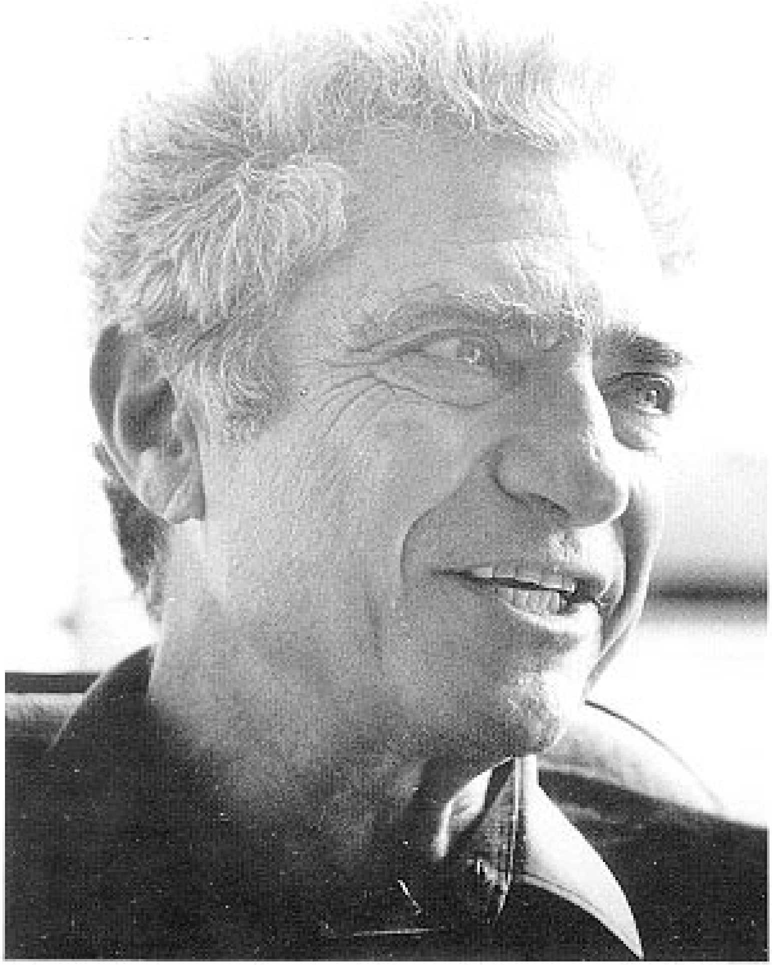


Figure 9. Spark distribution for 400 MeV/c electrons normalized to expected number of showers should be $\nu_\mu \neq \nu_e$. Also shown are the observed "shower" events.



Jack Steinberger

JACK STEINBERGER

I was born in Bad Kissingen (Franconia) in 1921. At that time my father, Ludwig, was 45 years old. He was one of twelve children of a rural `Viehhandler' (small-time cattle dealer). Since the age of eighteen he had been cantor and religious teacher for the little Jewish community, a job he still held when he emigrated in 1938. He had been a bachelor until he returned from four years of service in the German Army in the first World War. My mother was born in Nuremberg to a hop merchant, and was fifteen years the younger. Unusual for her time, she had the benefit of a college education and supplemented the meagre income with English and French lessons, mostly to the tourists which provided the economy of the spa. The childhood I shared with my two brothers was simple; Germany was living through the post-war depression.

Things took a dramatic turn when I was entering my teens. I remember Nazi election propaganda posters showing a hateful Jewish face with crooked nose, and the inscription “Die Juden sind unser Unglück”, as well as torchlight parades of SA storm troops singing "Wenn’s Juden Blut vom Messer fliesst, dann geht’s noch mal so gut”. In 1933, the Nazis came to power and the more systematic persecution of the Jews followed quickly. Laws were enacted which excluded Jewish children from higher education in public schools. When, in 1934, the American Jewish charities offered to find homes for 300 German refugee children, my father applied for my older brother and myself. We were on the SS Washington, bound for New York, Christmas 1934.

I owe the deepest gratitude to Barnett Faroll, the owner of a grain brokerage house on the Chicago Board of Trade, who took me into his house, parented my high-school education, and made it possible also for my parents and younger brother to come in 1938 and so to escape the holocaust. New Trier Township High School on the well-to-do Chicago North Shore, enjoyed a national reputation, and, with a swimming pool, athletic fields, cafeteria, as well as excellent teachers, offered horizons unimaginable to the young emigrant from a small German town.

The reunited family settled down in Chicago. We were helped to acquire a small delicatessen store which was the basis of a very marginal income, but we were used to a simple life, so this was no problem. I was able to continue my education for two years at the Armour Institute of Technology (now the Illinois Institute of Technology) where I studied chemical engineering. I was a good student, but these were the hard times of the depression, my scholarship came to an end, and it was necessary to work to supplement the family income.

The experience of trying to find a job as a twenty-year-old boy without connections was the most depressing I was ever to face. I tried to find any job in a chemical laboratory: I would present myself, fill out forms, and have the door closed hopelessly behind me. Finally through a benefactor of my older brother, I was accepted to wash chemical apparatus in a pharmaceutical laboratory, G.D. Seat-1 and Co., at eighteen dollars a week. In the evenings I studied chemistry at the University of Chicago, the weekends I helped in the family store.

The next year, with the help of a scholarship from the University of Chicago, I could again attend day classes, so that in 1942 I could finish an undergraduate degree in chemistry.

On 7 December 1942, Japan attacked the United States at Pearl Harbor. I joined the Army and was sent to the MIT radiation laboratory after a few months of introduction to electromagnetic wave theory in a special course, given for Army personnel at the University of Chicago. My only previous contact with physics had been the sophomore introductory course at Armour. The radiation laboratory was engaged in the development of radar bomb sights; I was assigned to the antenna group. Among the outstanding physicists in the laboratory were Ed Purcell and Julian Schwinger. The two years there offered me the opportunity to take some basic courses in physics.

After Germany surrendered in 1945, I spent some months on active duty in the Army, but was released after the Japanese surrender, to continue my studies at the University of Chicago. It was a wonderful atmosphere, both between professors and students and also among the students. The professors to whom I owe the greatest gratitude are Enrico Fermi, W. Zachariasen, Edward Teller and Gregor Wentzel. The courses of Fermi were gems of simplicity and clarity and he made a great effort to help us become good physicists also outside the regular class-room work, by arranging evening discussions on a widespread series of topics, where he also showed us how to solve problems. Fellow students included Yang, Lee, Goldberger, Rosenbluth, Garwin, Chamberlain, Wolfenstein and Chew. There was a marvellous collaboration, and I feel I learned as much from these fellow students as from the professors.

I would have preferred to do a theoretical thesis, but nothing within reach of my capabilities seemed to offer itself. Fermi then asked me to look into a problem raised in an experiment by Rossi and Sands on stopping cosmic-ray muons. They did not find the expected number of decays. After correcting for geometrical losses there was still a missing factor of two, and I suggested to Sands that this might be due to the fact that the decay electron had less energy than expected in the two-body decay, and that one might test this experimentally. When this idea was not followed, Fermi suggested that I do the experiment, instead of waiting for a theoretical topic to surface. The cosmic-ray experiment required less than a year from its conception to its conclusion, in the end of the summer of 1948. It showed that the muon's is a three-body decay, probably into an electron and two neutrinos, and helped lay the experimental foundation for the concept of a universal weak interaction.

There followed an interlude to try theory again at the Institute for Advanced Study in Princeton, where Oppenheimer had become director. It was a frustrating year: I was no match for Dyson and other young theoreticians assembled there. Towards the end I managed to find a piece of work I could do, on the decay of mesons via intermediate nucleons. I still remember how happy Oppenheimer was to see me come up with something, at last.

In 1949, Gian Carlo Wick, with whom I had done some work on the scattering of polarized neutrons in magnetized iron while still a graduate student at Chicago University, invited me to be his assistant at the University of California in Berkeley. There the experimental possibilities in the Radiation Laboratory, created by E.O. Lawrence, were so great that I reverted easily to my wild state, that is experimentation. During the year there, I had the magnificent opportunity of working on the just completed electron synchrotron of Ed McMillan. It enabled me to do the first experiments on the photoproduction of pions (with A.S. Bishop) to establish the existence of neutral pions (with W.K.H. Panofsky and J. Stellar) as well as to measure the pion mean life (with O. Chamberlain, R.F. Mozley and C. Weigand).

I survived only a year in Berkeley, partly because I declined to sign the anticommunist loyalty oath, and moved on to Columbia University in the summer of 1950. At its Nevis Laboratory, Columbia had just completed a 380 MeV cyclotron; this, for the first time, offered the possibility of experimenting with beams of π mesons. In the next years I exploited these beams to determine the spins and parities of charged and neutral pions, to measure the $\pi^-\pi^0$ mass difference and to study the scattering of charged pions. This work leaned heavily on the collaboration of Profs. D. Bodansky and A.M. Sachs, as well as of several Ph.D. students: R. Durbin, H. Loar, P. Lindenfeld, W. Chinowsky and S. Lokanathan.

These experiments all utilized small scintillator counters. In the early fifties, the bubble-chamber technique was discovered by Don Glaser, and in 1954 three graduate students, J. Leitner, N.P. Samios and M. Schwartz, and myself began to study this technique which had not as yet been exploited to do physics. Our first effort was a 10 cm diameter propane chamber. We made one substantial contribution to the technique, that was the realization of a fast recompression (within ~ 10 ms), so that the bubbles were recompressed before they could grow large and move to the top. This permitted chamber operation at a useful cycling rate. The first bubble-chamber paper to be published was from our experiment at the newly built Brookhaven Cosmotron, using a 15 cm propane chamber without magnetic field. It yielded a number of results on the properties of the new unstable (strange) particles at a previously unattainable level, and so dramatically demonstrated the power of the new technique which was to dominate particle physics for the next dozen years. Only a few months later we published our findings on three events of the type $\Sigma^0 \rightarrow \Lambda^0 + \gamma$, which demonstrated the existence of the Σ^0 hyperon and gave a measure of its mass. This experiment used a new propane chamber, eight times larger in volume, and with a magnetic field. This chamber also introduced the use of more than two stereo cameras, a development which is crucial for the rapid, computerized analysis of events, and has been incorporated into all subsequent bubble chambers.

In the decade which followed, the same collaborators, together with Profs. Plano, Baltay, Franzini, Colley and Prodell, and a number of new students, constructed three more bubble chambers: a 12” H_2 chamber as well as 30” propane and H_2 chambers, developed the analysis techniques, and performed a series of experiments to clarify the properties of the new particles. The experiments I remember with the most pleasure are:

the demonstration of parity violation in A decay, 1957;
the demonstration of the β decay of the pion, 1958;
the determination of the π^0 parity on the basis of angular correlation in the double internal conversion of the γ rays, 1962;
the determination of the ω and φ decay widths (lifetimes), 1962;
the determination of the Σ^0 - Λ^0 relative parity, 1963;
the demonstration of the validity of the AS = AQ rule in K^0 and in hyperon decays, 1964.

This long chain of bubble-chamber experiments, in which I also enjoyed and appreciated the collaboration of two Italian groups, the Bologna group of G. Puppi and the Pisa group of M. Conversi, was interrupted in 1961, in order to perform, at the suggestion of Mel Schwartz, and with G. Danby, J.M. Gaillard, D. Goulianos, L. Lederman and N. Mistri, the first experiment using a high-energy neutrino beam now recognized by the Nobel Prize, and described in the paper of M. Schwartz.

In 1964, CP violation was discovered by Christensen, Cronin, Fitch and Turlay. Soon after I found myself on sabbatical leave at CERN, and proposed, together with Rubbia and others, to look for the interference between K_S^0 and K_L^0 amplitudes in the time dependence of K^0 decay. Such interference was expected in the CP violation explanation of the results of Christensen et al., but not in other explanations which had also been proposed. The experiment was successful, and marked the beginning of a set of experiments to learn more about CP violation, which was to last a decade. The next result was the observation of the small, CP-violating, charge asymmetry in K_L^0 leptonic decay, in 1966. Measurement of the time dependence of this charge asymmetry, following a regenerator, permitted a determination of the regeneration phase; this, together with the earlier interference experiments, yielded, for the first time, the CP-violating phase $\varphi\eta_{+-}$ and, in consequence, as well as the observed magnitudes of the CP-violating amplitudes in the two-pion and the leptonic decays, certain checks of the superweak model. The same experiment also gave a more sensitive check of the AS = AQ rule, an ingredient of the present Standard Model.

In 1968, I joined CERN. Charpak had just invented proportional wire chambers, and this development offered a much more powerful way to study the K^0 decay to which I had become addicted. Two identical detectors were constructed, one at CERN together with Filthuth, Kleinknecht, Wahl, and others, and one at Columbia together with Christensen, Nygren, Carithers and students. The Columbia beam was long, and therefore contained no K_S but only K_L, the CERN beam was short, and therefore contained a mixture of K_S and K_L. It was contaminated by a large flux of A", and so was also a hyperon beam, permitting the first measurements of A" cross-sections as well as the Coulomb excitation of Λ^0 to Σ^0, a difficult and interesting experiment carried out chiefly by Steffen and Dydak. The most important result to come from the Columbia experiment was the observation of the rare decay $K_L \rightarrow \mu^+\mu^-$ with a branching ratio compatible with theoretical predictions based on unitarity. Previously, a Berkeley experiment had searched in vain for this decay and had claimed an upper limit in violation of unitarity. Since unitarity is fundamental to field theory, this result had a certain importance.

The CERN experiment, which extended until 1976, produced a series of precise measurements on the interference of K_S and K_L, in the two-pion and leptonic decay modes, thus leading us to obtain highly precise results on the CP-violating parameters in K^0 decay. I believe the experiment was beautiful, and take some pride in it, but the results were all in agreement with the super-weak model and so did little towards understanding the origin of CP violation.

In 1972, the K^0 collaboration of CERN, Dortmund and Heidelberg was joined by a group from Saclay, under R. Turlay, to study the possibilities for a neutrino experiment at the CERN SPS then under construction. The CDHS detector, a modular array of magnetized iron disks, scintillation counters and drift chambers, 3.75 m in diameter, 20 m long, and weighing 1200 t, was designed, constructed, and exposed to different neutrino beams at the SPS during the period 1977 to 1983. It provided a large body of data on the charged-current and neutral-current inclusive reactions in iron, which permitted first of all the clearing away of a number of incorrect results, e.g. the "high-y anomaly" produced at Fermilab, allowed the first precise and correct determination of the Weinberg angle, demonstrated the existence of right-handed neutral currents, provided measurements of the structure functions which gave quantitative support to the quark constituent model of the nucleon, and, through the Q^2 evolution of the structure functions, gave quantitative support to QCD. The study of multimuon events gave quantitative support to the GIM model of the Cabibbo current through its predictions on charm production.

In the CDHS experiment we were about thirty physicists. Since 1983, I have been spokesman for a collaboration of 400 physicists engaged in the design and construction of a detector for the 100 + 100 GeV e^+e^- Collider, LEP, to be ready at CERN in the beginning of 1989. In the meantime I had also helped to design an experiment to compare CP violation in the charged and neutral two-pion decay of the K_L^0. This experiment was the first to show "direct" CP violation, an important step towards the understanding of CP violation.

In 1986, I retired from CERN and became part-time Professor at the Scuola Normale Superiore in Pisa. However, my chief activity continues as before in my research at CERN.

I am married to Cynthia Alff, my former student and now biologist, and we have two marvelous children, Julia, 14 years old, and John, 11 years old. From an earlier marriage to Joan Beauregard, there are two fine sons, Joseph Ludwig and Richard Ned.

I play the flute, unfortunately not very well, and have enjoyed tennis, mountaineering and sailing, passionately.

EXPERIMENTS WITH HIGH-ENERGY NEUTRINO BEAMS

Nobel Lecture, December 8, 1988

by

JACK STEINBERGER

CERN, Geneva, Switzerland, and Scuola Normale Superiore, Pisa, Italy

1. INTRODUCTION

High-energy neutrino beams have found intensive and varied application in particle physics experimentation in the last decades. This review is constrained to a few of the most fruitful examples: the discovery of neutral currents, the measurement of the Weinberg angle, the study of weak currents and the consequent test of the electroweak theory, the study of nucleon quark structure, and the testing of quantum chromodynamics (QCD). Other studies such as the production of "prompt" neutrinos, the search for finite neutrino masses and neutrino oscillations, the search for heavy leptons or other new particles, or the measurement of proton and neutron structure functions, elastic and pseudoelastic cross-sections, and other exclusive processes, are not discussed here. Neutrino experiments have been pursued vigorously at the Brookhaven National Laboratory, at Fermilab, and at CERN. It is fair to say that they have made large contributions to our understanding of particle physics.

2. NEUTRINO BEAMS

Present neutrino beams are produced in four steps: i) production of secondary hadrons in the collision of high-energy protons on a fixed target; ii) momentum (charge) selection and focusing of the hadrons; iii) passage of the beam through an (evacuated) decay region, long enough to permit a substantial fraction of the hadrons to decay; iv) absorption of the remaining hadrons and the muons that are produced along with the 'neutrinos in a shield of adequate thickness. The two-body decays $\pi^{+(-)}\rightarrow\mu^{+(-)}+\nu(\bar{\nu})$ and $K^{+(-)}\rightarrow\mu^{+(-)}+\nu(\bar{\nu})$ account for ~ 97% of the neutrino flux in present beams. Positive hadrons produce neutrinos, negative hadrons produce antineutrinos. Figures la and lb give an impression of the two hadron beam-forming options that are available, side by side, at CERN: a conventional, so-called narrow-band beam (NBB), and an achromatic, Van der Meer horn-focused, wide-band beam (WBB). The neutrino spectra produced by these two beams are very different, as shown in Figure 2. The

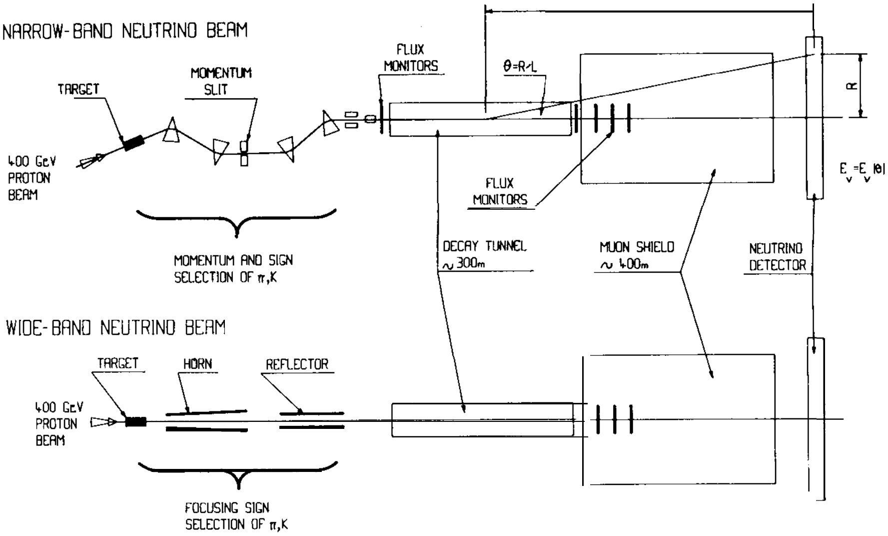


Figure *la* Sketch of narrow-band and wide-band neutrino beam layouts at CERN, showing disposition of primary target, focusing elements, decay region, shielding, and monitoring devices.

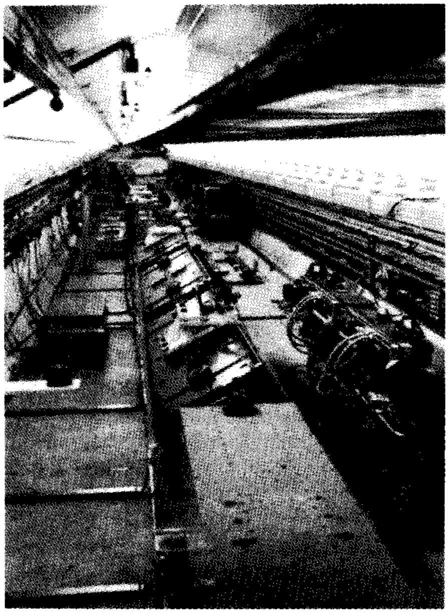


Figure *lb* View of the neutrino beam tunnel at the CERN SPS in 1976, before operations began. The NBB line is seen in the centre; on the right is the pulse transformer for the WBB horn, but the horn itself, destined for the pedestal on the left, is not yet installed. At the far end, the 2.5 m diameter titanium window of the evacuated decay region can be seen.

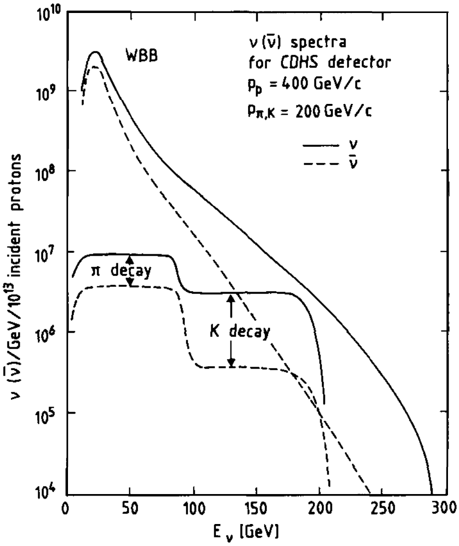


Figure 2 Neutrino and antineutrino energy spectra, calculated for the horn-focused WBB and the more conventional NBB.

WBBs are characterized by high intensity, a steep (generally undesirable) energy fall-off, and a substantial contamination of wrong-"sign" neutrinos. The NBBs have lower intensity, a flat energy dependence in the contribution from each of the two decays, and small wrong-sign background. They also have the important feature that the energy of the neutrino can be known, subject to a twofold π-K dichotomy, if the decay angle is known. In general this can be inferred from the impact parameter of the event in the detector.

3. DETECTORS

The low cross-sections of neutrinos are reflected in two general features of neutrino detectors:

i) they are massive;

ii) the target serves also as detector.

In the seventies, the most successful detectors were large bubble chambers. The most splendid of these were the cryogenic devices built at CERN and Fermilab, each with a volume of ~ 15 m^3, in large magnetic fields, and capable of operating with liquid hydrogen, deuterium, or neon. A picture of a typical neutrino event in the CERN chamber is shown in Figure 3. It is an example of the "charged-current" (CC) reaction $\nu+N\rightarrow\mu^-+$hadrons. However, one of the major discoveries at CERN was made not in this but in a large Freon-filled bubble chamber, affectionately called Gargamelle. The active volume was a cylinder 4.8 m long and 1.9 m in diameter, for a volume of about 13 m^3, inside a magnet producing a field of 2 T. Figure 4 gives some impression of its size.

The bubble chamber has now been largely replaced by detectors based on electronic detection methods. As an example, I mention here the CDHS

Figure 3 A typical neutrino event as observed in the Big European Bubble Chamber (BEBC) filled with neon at the CERN 450 GeV Super Proton Synchrotron (SPS) accelerator. The muon can be seen on the left. It has been tagged by an external muon identifier. The many-particle hadron shower is to the right.

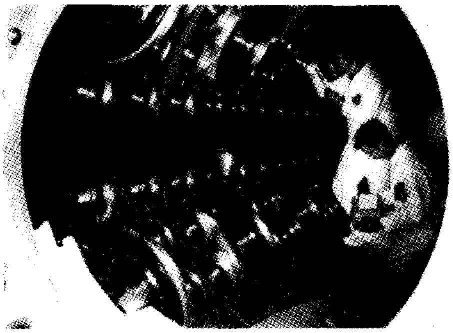


Figure 4 Preparation of the interior of the 13 m^3 bubble chamber Gargamelle, later to be filled with Freon. It is with this detector that the neutral currents were discovered.

(CERN-Dortmund-Heidelberg-Saclay Collaboration) detector used at CERN from 1977 to 1985. It consists of 19 modules made of iron plates 3.75 m in diameter, each with total iron thickness of 75 cm and a weight of ~ 65 t. The iron is toroidally magnetized to a field of 1.7 T by means of coils that pass through a hole in the centre.

Interleaved with the 5 cm thick iron plates are scintillator strips, which serve to measure the energy of the secondary hadrons by sampling the ionization. The typical hadron shower is ~ 25 cm in radius and ~ 1 m long, so the shower dimensions are very small compared with the size of the detector. The muon momenta are determined on the basis of curvature in the magnetic field, with the help of drift chambers inserted between the iron modules. These measure the positions of traversing tracks in three projections. The useful target weight is ~ 800 t. Figure 5 shows the CDHS experiment and Figure 6 a typical event of the same type as that shown in Figure 3.

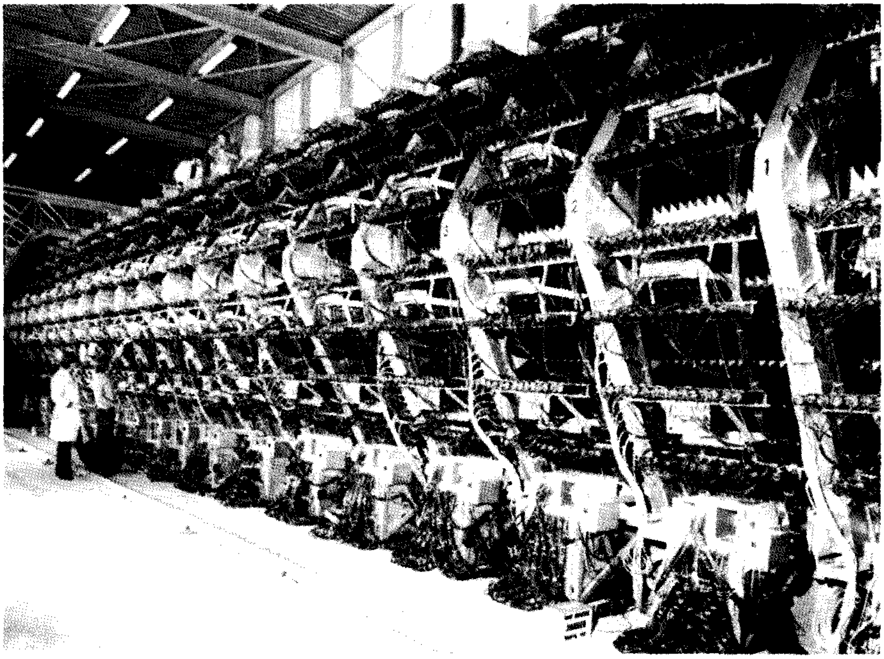


Figure 5 View along the 19 modules of the CDHS electronic neutrino detector at the SPS. The black light-guides and phototubes, which are used to measure the hadron energy, can be seen sticking out of the magnetized iron modules. The hexagonal aluminium structures are the drift chambers that measure the muon trajectories

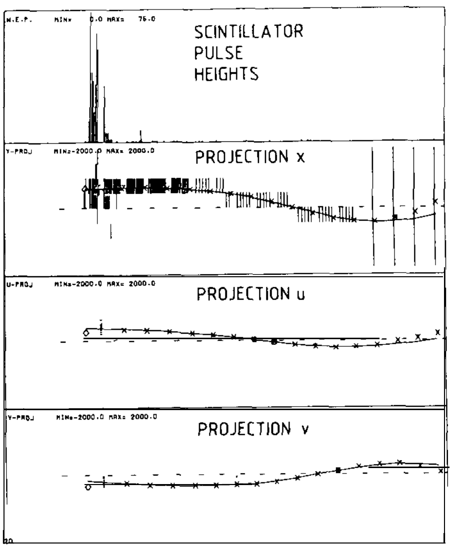


Figure 6 Computer reconstruction of a typical event of the reaction v + Fe$\to\mu^-$ + X. Four views are shown, with the horizontal axis along the beam direction. The top view shows the scintillator pulse height, or hadron energy, and its distribution along the detector. The next view shows the scintillatar hits as well as the horizontal wire hits and the reconstructed track in the x projection. The other two views show the wire hits and the reconstructed track for the ±60” projections.

4. NEUTRAL CURRENTS

4.1 Discovery

The evolution of the electroweak unified gauge theory in the late sixties and early seventies was a miraculous achievement, but one that had no immediate impact on the majority of particle physicists-certainly not on me-perhaps because it was a theoretical construct which left the existing experimental domain intact. However, it predicted some entirely new phenomena, and of these the neutral weak currents were the first to be discovered. The verification of neutral currents (NCs) established the theory overnight, and subsequent experiments on their detailed structure reinforced this. This observation' of neutral weak currents at CERN in 1973 by the Gargamelle group was the first great discovery made at CERN. It was followed 10 years later by the second-also a prediction of the same theory - the intermediate boson.

Figure 7 A "muonless" event in Gargamelle. All tracks stop or interact in the chamber. None could be a muon. The neutrino produces one K^+ and one K^0 meson. The K^+ meson interacts in the liquid and then decays. The invisible K^0 meson decays to two pions.

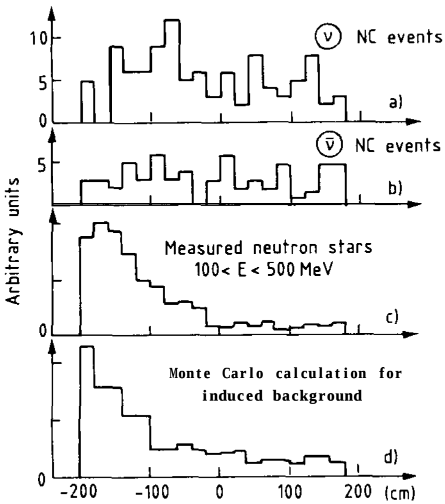


Figure 8 Distribution of the origin of muonless events along the beam direction in Gargamelle. Neutrino events are expected to be uniformly distributed, whereas neutron events should decrease with distance because of their absorption in the Freon. The nuclear mean free path in Freon is about 80 cm. The expected and observed distributions of neutron interactions are shown in the bottom two histograms. The muonless events are consistent with neutrino and inconsistent with neutron origin (Ref. 1).

The bubble chamber, built under the direction of A. Lagarrigue at the École Polytechnique in Paris, was exposed to neutrino and antineutrino WBBs at the CERN 24 GeV proton accelerator. The normal CC reactions,

$$\nu(\bar{\nu}) + \mathrm{N} \rightarrow \mu^-(\mu^+) + \text{hadrons},$$

were found as usual, but NC "muonless" reactions,

$$\nu(\bar{\nu}) + \mathrm{N} \rightarrow \nu(\bar{\nu}) + \text{hadrons},$$

which had hardly been looked for before-and therefore had not been found- were there as well. Such an event is shown in Figure 7. These events were selected on the basis of no muon candidate among the observed particles. The main experimental challenge was to show that they were not due to stray neutrons in the beam. I myself was a sceptic for a long time, and I lost a bottle or two of good wine on this matter. However, the neutron background would be expected to decrease exponentially along the length of the chamber, roughly with the neutron mean free path in Freon. Instead, the event distribution was flat, as expected for neutrino events (see Fig. 8.). I have never enjoyed paying up a debt more than at the dinner we gave for the winners, very good friends, Jacques Prentki, John Iliopoulos and Henri Epstein.

The ratios of the cross-sections

$$R_\nu = \frac{\sigma_\nu^{NC}}{\sigma_\nu^{CC}} \quad \text{and} \quad R_{\bar{\nu}} = \frac{\sigma_{\bar{\nu}}^{NC}}{\sigma_{\bar{\nu}}^{CC}}$$

are given in the electroweak theory in terms of the Weinberg angle θ_w:

$$R_\nu = \frac{1}{2} - \sin^2\theta_w + (1 + r)\frac{5}{9}\sin^4\theta_w \tag{1}$$

and

$$R_{\bar{\nu}} = \frac{1}{2} - \sin^2\theta_w + \left(1 + \frac{1}{r}\right)\frac{5}{9}\sin^4\theta_w\ , \tag{2}$$

where r is, the ratio of antineutrino to neutrino, CC total cross-sections: $r = \sigma^{CC,\bar{\nu}}/\sigma^{CC,\nu} = 0.48 \pm 0.02$ experimentally. On the basis of these ratios, the experiment yielded a first measure of $\sin^2\theta_w$, that was not very different from present, more precise determinations. In the same exposure a beautiful example of another NC process, the scattering of an antineutrino on an electron, was also found*.

4.2 Precision measurement of $\sin^2\theta_w$ and right-handed neutral currents

The higher energies that became available a few years later at Fermilab and CERN made the study of NC processes much easier. The muons of the CC background had now a greater penetration power, which permitted cleaner separation of NC and CC events. Also, with the advent of the higher energies, the advantage in the study of inclusive neutrino scattering had shifted to electronic detection techniques. In the period 1977 to 1985, hadronic NC neutrino scattering was studied extensively by the CDHS

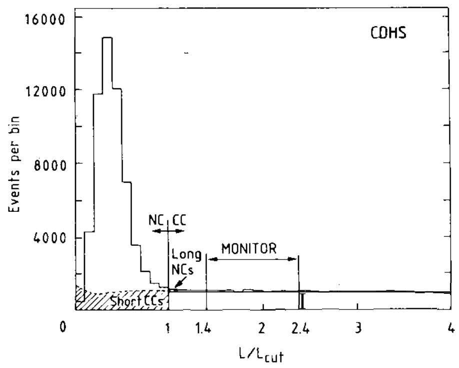


Figure 9 **Identification, by the CDHS Collaboration, of NC events by their short** event length. The peak at small event lengths is due to NC events. The long tail is due to muonic events, which must be subtracted under the peak to give the NC rate (Ref. 3).

Collaboration at CERN in order to get a more precise value for θ_w[3] and to check the prediction of the electroweak theory for the ratio of right-handed to left-handed NCs[4]. The NC events are selected on the basis of short event length, i.e. the short penetration of the hadronic shower compared with that of the muon of CC events. This is illustrated in Figure 9. A 15% background of CC events is subtracted. The neutrino NC-to-CC ratio $\mathbf{R}_\nu$ yielded the most precise value of the weak mixing angle available at present, $\sin^2\theta_w = 0.227 \pm 0.006$. Once $\sin^2\theta_w$ is known, the antineutrino ratio $\mathbf{R}_{\bar{\nu}}$ follows from Eq. (2). Its measurement provided a sensitive test of the electroweak theory, and confirmed it in its simplest form. The presence of right-handed NCs (CCs are purely left-handed) in the amount predicted by the theory could be demonstrated by comparing the hadron energy distributions of the NC and CC processes. The result is shown in Figure 10.

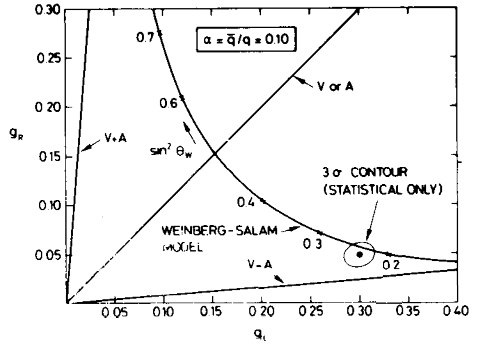


Figure 10 Strengths of the left- and right-handed neutral currents. If the NC were purely left-handed, as is the case for the CC, the experimental point would be expected to fall on the V-A line. The experiment shows a right-handed component, which is **just that expected in the electroweak theory** (Weinberg-Salam model) (Ref. 4).

4.3 Neutrino-electron scattering

The elastic-scattering reactions of neutrinos on atomic electrons

$$\nu_\mu + e^- \rightarrow \nu_\mu + e^- \qquad \text{and} \qquad \bar{\nu}_\mu + e^- \rightarrow \bar{\nu}_\mu + e^-$$

proceed via NCs. They are characterized by small cross-sections-smaller than their hadronic counterparts by the mass ratio m_e/m_p because of the smaller c.m. energies-and, for the same reason, by small electron production angles, $\theta_e \approx \sqrt{m_e/E_\nu}$. Until now, these angles have not been resolved by the experiments, so only total cross-sections have been measured. The expectations in the electroweak theory are:

$$\sigma^{\nu,e} = \frac{G_F^2\, E m_e}{\pi}\left(1 - 4\sin^2\theta_w + \frac{16}{3}\sin^4\theta_w\right)$$

and

$$\sigma^{\bar{\nu},e} = \frac{G_F^2\, E m_e}{\pi}\left(\frac{1}{3} - \frac{4}{3}\sin^2\theta_w + \frac{16}{3}\sin^4\theta_w\right) .$$

These reactions can also serve to test this theory, and have the advantage that strongly interacting particles are not involved, so that the understanding of strong-interaction corrections is not necessary in the interpretation. They have the experimental disadvantage of low rates and consequent large background. The best results at present are from a BNL experiment[5] using relatively low energy neutrinos, $E \approx 1.5$ GeV, and a 140 t detector entirely composed of many layers of plastic scintillator and drift chambers. The background is subtracted on the basis of the distribution in the production angle of the electron shower (see Figure 11). Instead of comparing the neutrino and antineutrino cross-sections directly with the theory, the authors form the ratio of the two, which is less sensitive to some systematic errors. From this they find the result $\sin^2\theta_w = 0.209 \pm 0.032$. A CERN group reports a similar result, with $\sin^2\theta_w = 0.211 \pm 0.037$. The agreement with other methods of obtaining this angle is an important confirmation of the theory. A massive experiment to improve the precision is currently under way at CERN.

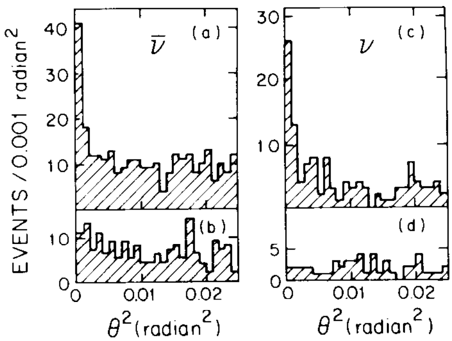


Figure 11 Identification of neutrino-electron events on the basis of their small angle with respect to the beam in the experiment of Ahrens et al.“. The peak at small angles in the two top graphs is due to neutrino-electron scattering. The bottom graphs show the flat distribution observed if photons rather than electrons are detected. This shows the angular distribution of background events.

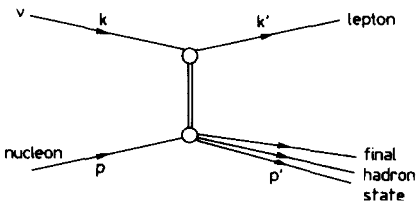
5. NEUTRINO-NUCLEON INCLUSIVE SCATTERING AND THE QUARK STRUCTURE OF HADRONS

5.1 Phenomenology

We consider the CC reactions,

$$\nu + N \rightarrow \mu^- + \text{hadrons} \quad \text{and} \quad \bar{\nu} + N \rightarrow \mu^+ + \text{hadrons},$$

independently of the final hadron configuration. This is called the inclusive process. It is assumed that the lepton vertex is described by the vector-axial vector current of the electroweak theory. Let k be the initial and k' the final lepton energy-momentum four vectors, p that of the incident nucleon, and p' that of the final hadron state:



Define the kinematic variables:

$Q^2 \equiv (k-k')^2 = 4EE' \sin^2\theta/2$,
$\nu \equiv p.Q^2/m_p = E_{h'} - m_{p'} \simeq E_h$
(the energy of the final-state hadrons in the laboratory system),
$x \equiv Q^2/2m_p\nu$, $0<x<1$,
$y \equiv \nu/E \simeq E_h/E$, $0<y<1$,

where E and E' are the energies of the initial neutrino and final muon, respectively, and θ is the angle between these, all in the laboratory system. The cross-sections can be written in terms of three structure functions, each a function of the variables x and Q^2 that characterize the hadronic vertex:

$$\frac{d^2\sigma^{\nu(\bar{\nu})}}{dxdy} = \frac{G^2 m_p E_\nu}{2\pi}\{F_2(x,Q^2)[1+(1-y)^2] - y^2F_L(x,Q^2) \underset{(-)}{+} xF_3(x,Q^2)[1-(1-y)^2]\} .$$

The functions $F_2(x,Q^2)$, $xF_3(x,Q^2)$, and $F_L(x,Q^2)$ are the three structure functions that express what happens at the hadron vertex. The sum of neutrino and antineutrino cross-sections has the same structure-function dependence as does the cross-section for charged leptons:

$$\frac{d^2\sigma^{\ell^\pm}}{dxdy} = \frac{2\pi\alpha^2 m_p E}{Q^4}\{F_2^{\ell^\pm}(x,Q^2)[1+(1-y)^2] - y^2F_L(x,Q^2)\} .$$

5.2 Quark structure of the nucleon

In 1969, at the newly completed S-mile linear electron accelerator at SLAC, it was discovered' that in electron-proton collisions, at high momentum transfer, the form factors were independent of Q^2. This so-called "scaling" behaviour is characteristic of "point", or structureless, particles. The interpretation in terms of a composite structure of the protons- that is, protons composed of point-like quarks -was given by Bjorken[8] and Feynman[9].

Neutrinos are projectiles *par excellence* for investigating this structure, in part because of the heavy mass of the intermediate boson, and in part because quarks and antiquarks are scattered differently by neutrinos owing to the V-A character of the weak currents; they can therefore be distinguished in neutrino scattering, whereas in charged-lepton scattering this is not possible. The quark model makes definite predictions for neutrino-hadron scattering, which are beautifully confirmed experimentally. Many of the predictions rest on the fact that now the kinematical variable x takes on a physical meaning: it can be interpreted as the fraction of the nucleon momentum or mass carried by the quark on which the scattering takes place. The neutrino experiments we review here have primarily used iron as the target material. Iron has roughly equal numbers of protons and neutrons. For such nuclei, the cross-sections can be expressed in terms of the total quark and total antiquark distributions in the proton. Let u(x), d(x), s(x), c(x), etc., be the up, down, strange, charm, etc., quark distributions in the proton. The proton contains three "valence" quarks: two up-quarks and one down-quark. In addition, it contains a "sea" of virtual quark-antiquark pairs. The up valence-quark distribution is $u(x) - \bar{u}(x)$, and the down valence-quark distribution is $d(x) - \bar{d}(x)$. The sea quarks and antiquarks have necessarily identical distributions, so that $s(x) = \bar{s}(x)$, $c(x) = \bar{c}(x)$, etc. For the neutron, u and d change roles, but s and c are the same. Let

$$q(x) = u(x) + d(x) + s(x) + c(x) + \dots$$

and

$$\bar{q}(x) = \bar{u}(x) + \bar{d}(x) + \bar{s}(x) + \bar{c}(x) + \dots$$

be the total quark and antiquark distributions of the proton, respectively. For spin-1/2 quarks interacting according to the Standard Model, for a target with equal numbers of protons and neutrons, and for

$Q^2 \ll m_W^2$ and $m_p \ll E$:

$$\frac{d^2\sigma^\nu}{dxdy} = \frac{G_F^2 Em_p}{\pi} x[q(x) + (1-y)^2\bar{q}(x)]$$

and

$$\frac{d^2\sigma^{\bar{\nu}}}{dxdy} = \frac{G_F^2 Em_p}{\pi} x[\bar{q}(x) + (1-y)^2 q(x)]$$

Comparison with the expression for the cross-section in terms of structure functions then gives these functions in terms of quark distributions:

$F_2(x,Q^2) = x[q(x) + \bar{q}(x)]$: $q(x) + \bar{q}(x)$ is the total quark + antiquark distribution;

$xF_3(x,Q^2) = x[q(x) - \bar{q}(x)]$: $q(x) - \bar{q}(x)$ is the 'valence'-quark distribution;

$F_L(x,Q^2) = 0$: this is a consequence of the spin-1/2 nature of the quarks.

From these simple expressions for the cross-sections, in terms of quark structure, several tests of the quark model are derived. For the experimental comparisons, we take the CDHS experiments[10]. It should be noted that the measurements in the detector, i.e. the hadron energy and the muon momentum, are just sufficient to define the inclusive process.

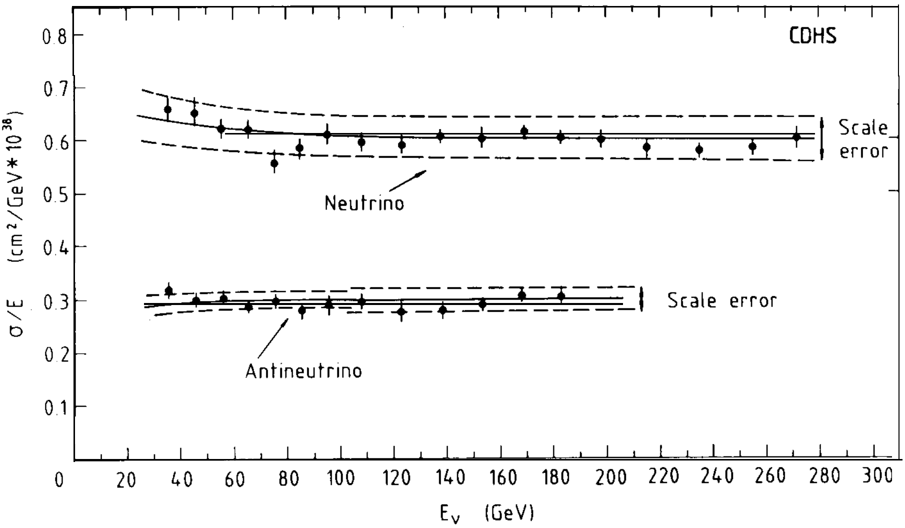


Figure 12 Total neutrino and antineutrino cross-sections per nucleon divided by neutrino energy. The flat 'scaling' behaviour is a consequence of the point-like interaction of the constituents (Ref. 10).

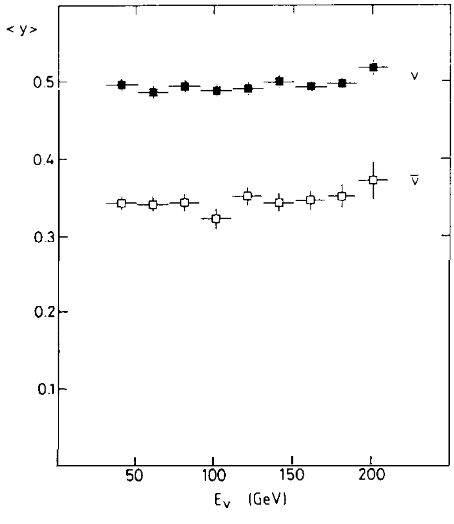


Figure 13 The average of y (the fraction of the neutrino energy transmitted to the final hadron state) as a function of the neutrino energy, for neutrinos and antineutrinos. The uniformity is a consequence of scaling, which in turn is a consequence of the point-like interaction of the quark (Ref. 10).

1) Scaling. The independence of the differential cross-sections with respect to Q^2 is evident everywhere, over a large domain in Q^2. As one example, Figure 12 shows the linearity of the total cross-sections with neutrino energy; as another, Figure 13 shows the uniformity of the average of y with respect to neutrino energy; both examples are consequences of scaling. Small deviations from scaling are observed in the structure functions, as we will see later, but these have their explanation in the strong interactions of the quarks.

2) The y-dependence of the cross-sections. We expect

$$\frac{d\sigma^{\nu}}{dy} + \frac{d\sigma^{\bar{\nu}}}{dy} \propto [1 + (1 - y)^2] \int x[q(x) + \bar{q}(x)]\, dx$$

and

$$\frac{d\sigma^{\nu}}{dy} - \frac{d\sigma^{\bar{\nu}}}{dy} \propto [1 - (1 - y)^2] \int x[q(x) - \bar{q}(x)]\, dx \ .$$

The agreement with this expectation is quite good, as can be seen from Figure. 14. A corollary of this agreement is that $F_L(x)$ is small. It is found that $\int F_L(x)dx/\int F_2(x)dx \simeq 0.1$. Again, this deviation from the simple quark picture is understood in terms of the strong interactions of the quarks, as we will see later.

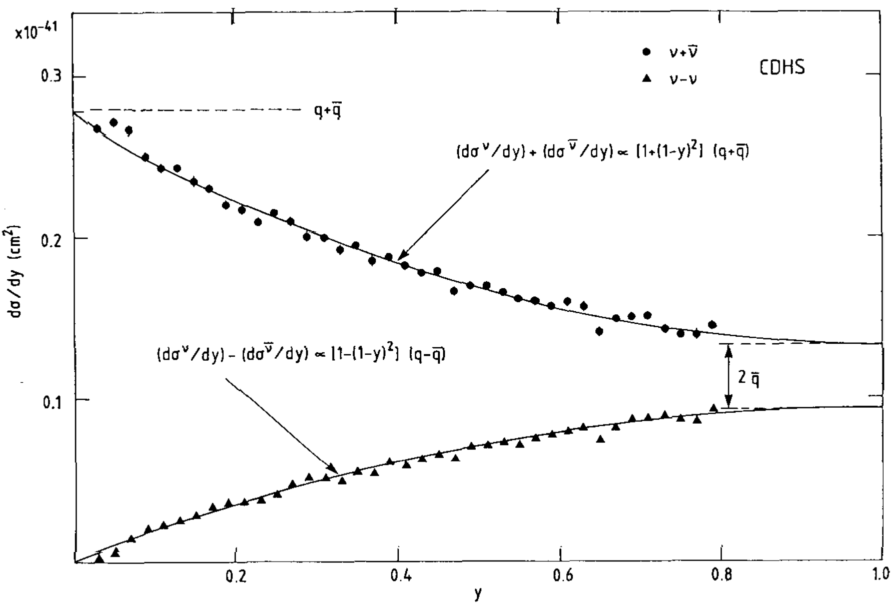


Figure 14 The y-dependence of the sum and the difference of neutrino and antineutrino cross-sections. Spin-$^1/_2$ quarks are expected to have y-dependences $1 + (1-y)^2$ for the sum and $1 - (1-y)^2$ for the difference (Ref. 10).

3) Correspondence between $F_2^{\ell\pm}(y)$ and $F_2^{\nu}(x)$. Both are proportional to q(x) + q(x), and so are expected to have the same x-dependence in the simple quark model. They are related by the factor

$$\frac{F_2^{\ell\pm}(x)}{F_2^{\nu}(x)} = \frac{1}{2}\left[\left(\frac{2}{3}\right)^2 + \left(-\frac{1}{3}\right)^2\right] = \frac{5}{18}$$

Here $^2/_3$ and $-^1/_3$ are the up- and down-quark electric charges, respectively. The agreement in shape and magnitude, shown in Figure 15, not only supports the quark picture, but also demonstrates the third integral quark electric charges.

4) $\int xF_3(x)\,dx/x = 3$. Since $xF_3(x) = x[q(x) - \bar{q}(x)]$ in the quark model and $q(x) - \bar{q}(x)$ is the valence-quark distribution, this sum rule states that there are three valence quarks in the nucleon. The experimental demonstration is not without problems, because the ν and $\bar{\nu}$ cross-sections are finite as $x \rightarrow 0$, and the difference, which is $xF_3(x)$, has a consequent large error at small x, which is divided by x as $x \rightarrow 0$. However, all experiments give a value near 3, with typical uncertainties of $\sim$ 10%.

Together with the charged-lepton inclusive scattering experiments, the neutrino experiments leave no doubt about the validity of the quark picture of nucleon structure. In addition, the neutrino experiments are unique in offering the possibility of measuring independently the quark and antiquark distributions in the nucleon, shown in Figure 15.

If the quarks were the sole nucleon constituents, we would expect $\int F_2\,dx = \int x[q(x) + q(x)]\,dx$ to be equal to 1. Experimentally, $\int F_2(x)\,dx = 0.48 \pm 0.02$. We should have expected that some of the nucleon momentum is carried by the gluons, the mesons that bind the quarks. The experimental result is therefore interpreted to mean that gluons account for about half of the nucleon momentum (or mass).

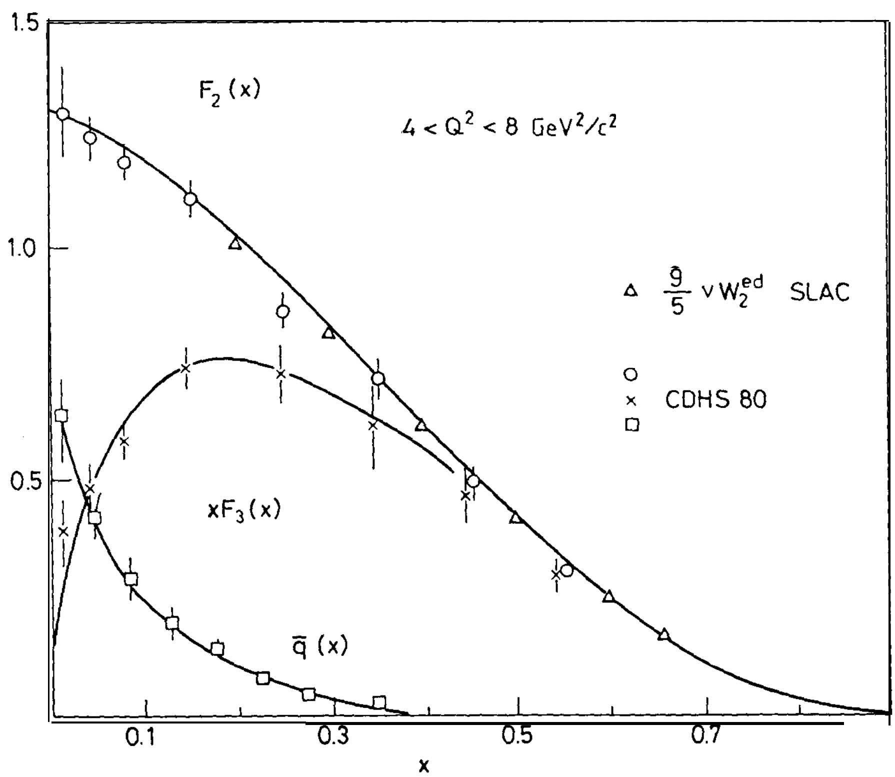


Figure 15 The structure functions $xF_3(x)$, $F_2(x)$, and $\bar{q}(x)$. In the simple quark picture $F_2(x) = x[q(x) + \bar{q}(x)]$ and $xF_3(x) = x[q(x) - \bar{q}(x)]$.

5.3 *Neutrino* scattering and quantum *chromodynamics (QCD)*

QCD is the elegant new gauge theory of the interaction of quarks and gluons, which describes the binding of quarks into the hadrons. Deep-inelastic lepton scattering provided a means of testing the predictions of this important theory and gave it its first experimental support. So far, no one has succeeded in calculating low-energy hadronic phenomena such as the wave functions of quarks in hadrons, because of the large coupling constant that frustrates perturbation methods at low energy. At high Q^2, however, the effective coupling constant becomes logarithmically smaller, and perturbation calculations become credible. The theory predicts 'scaling violations' in the form of a 'shrinking' of the structure functions towards smaller x as Q^2 gets larger. This is observed experimentally, as can be seen from Figure 16. In the theory, the 'shrinking' is the consequence of the

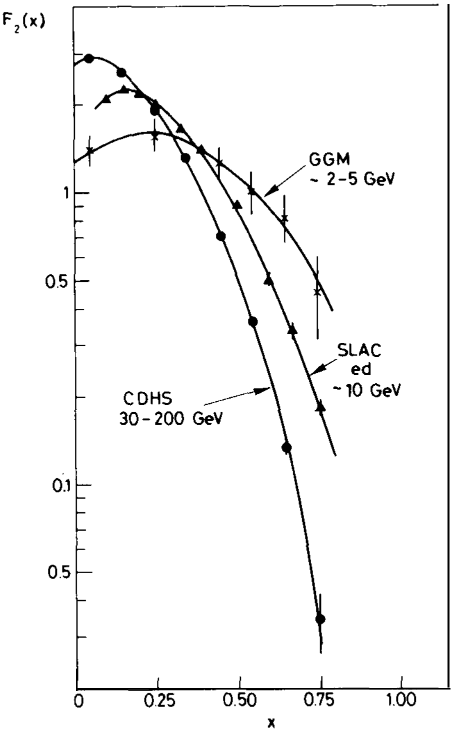


Figure 16 Scaling is only approximately true for the structure functions. Early measurements of $F_2(x)$ in three different energy domains exhibit shrinking, as expected in the QCD theory.

emission of gluons in the scattering process. This emission can be calculated. The Q^2 evolution at sufficiently high Q^2 is therefore quantitatively predicted by the theory. In neutrino experiments, this Q^2 evolution could be measured, and these measurements confirmed the theory and contributed to its acceptance. In the case of xF_3, the theoretical predictions have only one free parameter, the coupling constant α_S. In the case of F_2, the Q^2 evolution is coupled to the gluon distribution $G(x,Q^2)$. The experimental Q^2 evolutions of xF_3 and F_2 in the latest CDHS experiment are shown, together with their QCD fits, in Figure 17 and 18. The theory fits the data adequately. These fits give a value for the parameter A in the running strong-coupling constant,

$$\alpha_s = [6/(33 - 2N_f) \ln (Q^2/\Lambda^2)],$$

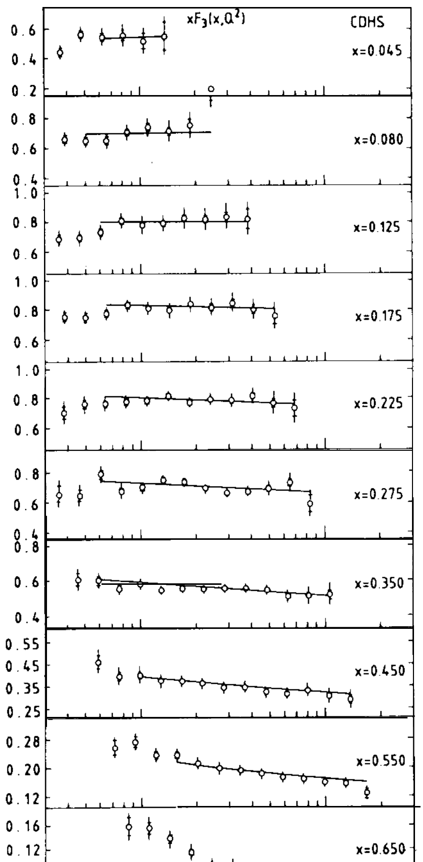


Figure 17 Variation of $xF_3(x,Q^2)$ with In Q^2 in different x bins. The Q^2 evolution predictions of QCD with A = 128 MeV are also shown (Ref. 10).

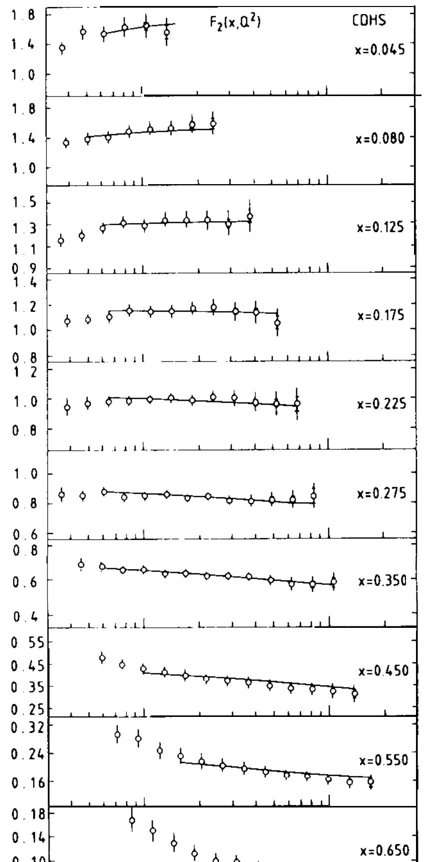


Figure 18 Variation of $F_2(x,Q^2)$ with In Q^2. The QCD fit is also shown.

where N_f = number of excited quark flavours ($N_f \simeq 4$ in this experiment), A $\simeq$ 100 MeV. They also give the gluon distribution shown in Figure 19. These QCD comparisons suffer somewhat from the fact that Q^2 is still too low to reduce non-perturbative effects to a negligible level, but the calculable perturbative effects dominate and are confirmed by the experiments. Perturbative QCD also predicts a non-zero longitudinal structure function $F_L(x,Q^2)$ as another consequence of the emission of gluons. This prediction is compared with the CDHS experimental results in Figure 20. Again, the experiment lends support to the theory.

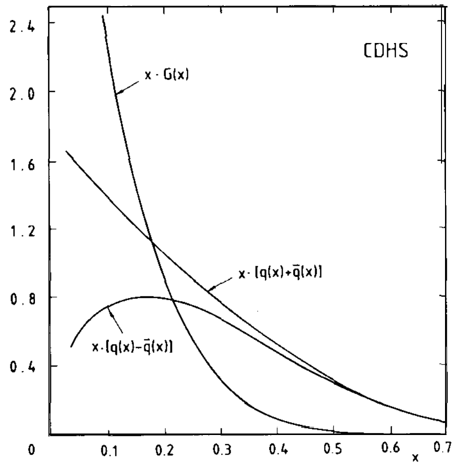


Figure 19 The gluon distribution G(x) derived from the QCD fits to $F_2(x,Q^2)$, $\bar{q}(x,Q^2)$, and $xF_3(x,Q^2)$ (Ref. 10).

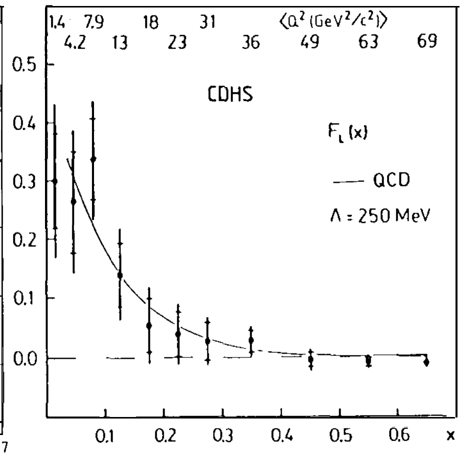


Figure 20 The structure function $F_L(x)$ associated with longitudinally polarized intermediate bosom, and the QCD predictions. In the simple quark model, F_L is zero (Ref. 10).

6. NEUTRINO INTERACTIONS, THE GIM WEAK CURRENT, AND THE STRANGE QUARK IN THE NUCLEON

Among the most beautiful results obtained with neutrino beam experiments are those concerning the opposite-sign `dimuons' first observed at Fermilab” and studied in detail in the CDHS experiments”. These reactions occur with roughly 1/100 of the rate of the dominant single-muon events. The experiments are interesting, on the one hand because they confirm the doublet structure of the quark weak current proposed some years ago by Glashow, Iliopoulos and Maiani[13], and which is fundamental to the electroweak theory, and on the other hand because they give such a vivid confirmation of the nucleon quark structure altogether.

The origin of the extra muon was quickly understood as being due to the production of charmed quarks and their subsequent muonic decay. In the GIM model, the charm-producing reactions are

	GIM cross-section proportionality	
$\nu + d \rightarrow \mu^- + c;\ c \rightarrow \mu^+ + \ldots$	$xd(x)\sin^2\theta_C$	(3)
$\nu + s \rightarrow \mu^- + c;\ c \rightarrow \mu^+ + \ldots$	$xs(x)\cos^2\theta_C$	(4)

and

$\bar{\nu} + \bar{d} \rightarrow \mu^+ + \bar{c};\ \bar{c} \rightarrow \mu^- + \ldots$	$x\bar{d}(x)\sin^2\theta_C$	(5)
$\bar{\nu} + \bar{s} \rightarrow \mu^+ + \bar{c};\ \bar{c} \rightarrow \mu^- + \ldots$	$x\bar{s}(x)\cos^2\theta_C$	(6)

The identification of the extra muon events with charm decay is experimentally confirmed in a number of ways:

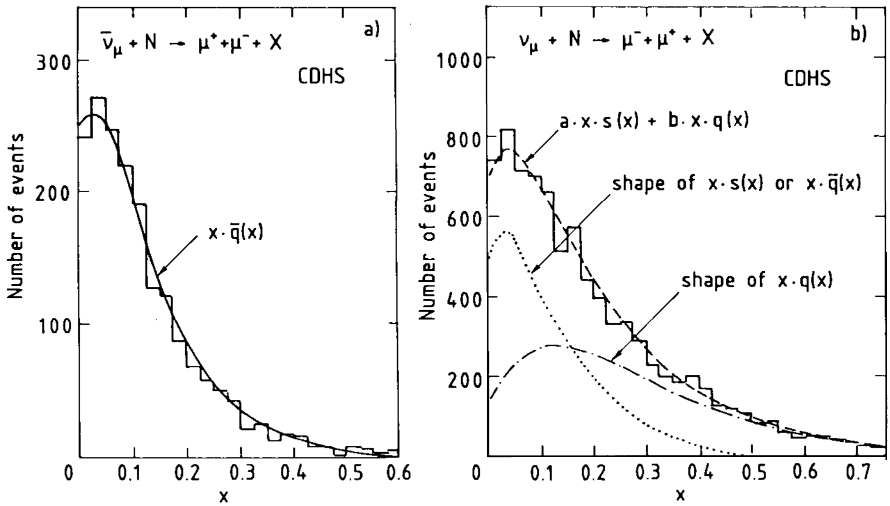


Figure 21 The x-distributions of opposite-sign dimuon events. a) For antineutrinos. The dominant process is $\bar{\nu} + \bar{s} \to \mu^+ + \bar{c}$. The observed x-distribution is therefore that of the strange sea in the nucleon.
b) For neutrinos. The process is v + s or d - μ^- + c. The shape allows the determination of the relative contributions of s and d quarks, and therefore the relative coupling constant. This confirmed the GIM prediction (Ref. 12).

i) opposite-sign muons are produced, like-sign ones are not;
ii) in general, the extra muon has little energy;
iii) the extra muon is correlated, as expected, to the direction of the hadron shower, of which the charmed particle is a part.

The GIM paper’” preceded the experimental discovery of charm by five years. It was proposed because of the theoretical attractiveness of the doublet structure of the weak currents. The predictions were precise. The cross-sections are proportional to $\sin^2\theta_C$ for d and $\bar{d}$ quarks and to $\cos^2\theta_C$ for s and $\bar{s}$ quarks. The Cabibbo angle θ_C was previously known, with $\cos^2\theta_C$ = 0.97, close to 1, and $\sin^2\theta_C$ = 0.05, very much smaller. Reactions (3) and (4), or (5) and (6), are not experimentally separable since the target nucleon contains both sand d quarks, and the final state is the same. In the antineutrino case, reaction (6) dominates (5) because $\sin^2\theta_C$ is so small. For each event, x and y are measured as for single-muon events. Therefore, the x-distribution for antineutrino dimuon production, shown in Figure 21a, measures the amount and the shape of the strange sea s(x).

In the neutrino reactions, the smallness of $\sin^2\theta_C$ for reaction (3) is very closely compensated by the fact that d(x), containing also valence quarks, is much greater than s(x) of reaction (4). By fitting, it can be seen that the x-distribution in Figure 21 b is a roughly equal mixture of s(x) as obtained with the antineutrinos and d(x), previously known from the normal CC reactions. The ratio of the two contributions is a measure of θ_C as it enters the charm production reaction. The Cabibbo angle obtained in this way is found to be equal, within errors, to θ_C measured in strange decays, as proposed in the

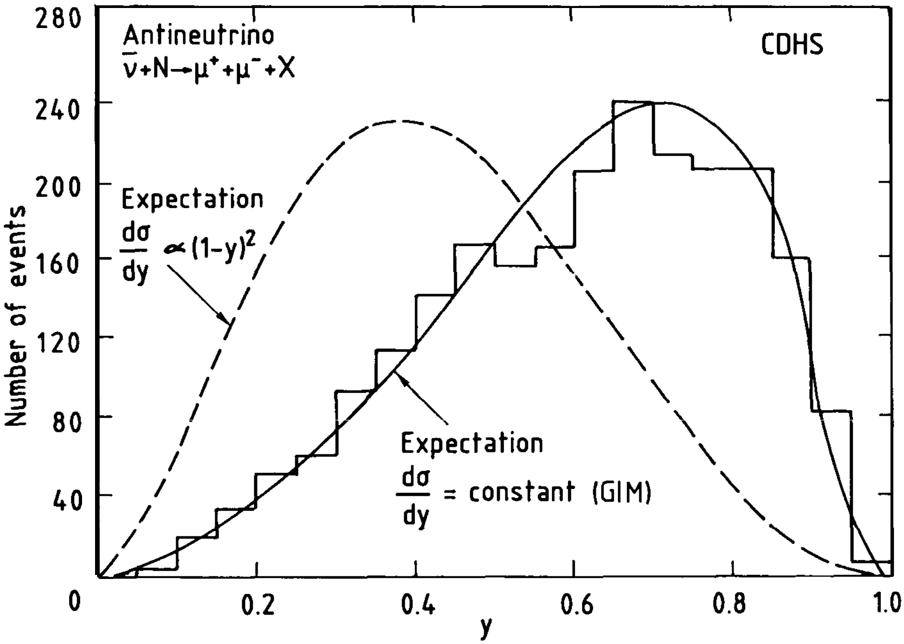


Figure 22 The y-distribution of $\bar{\nu}$-produced dimuons. The acceptance over the y-domain is unfortunately very non-uniform, because of the 5 GeV minimum energy required of each muon. The observed y-distribution agrees with an acceptance-corrected flat y-distribution as predicted by the GIM current, but differs strikingly from the $(1-y)^2$ distribution characteristic of the single-muon antineutrino cross-section (Ref. 12).

GIM hypothesis. Further support of the GIM current is provided by the y-distributions. They reflect the relative helicities of the neutrino and the struck quark: if the two helicities are the same, as is the case for all four charm-producing reactions, the expected y-distribution is flat; if they are opposite, as is the case for instance for $\nu + \bar{q}$ and $\bar{\nu} + q$, the expected distribution is $(1 - y)^2$. Both neutrino and antineutrino single-muon reactions are mixtures of the two, as we saw in Figure 14. The contrast is especially strong for antineutrinos, where the experimental single-muon y-distribution is dominated by $(1 - y)^2$, whereas the dimuon distribution is flat, as shown in Figure 22, again confirming the GIM picture.

7. CONCLUDING REMARKS

I have given some examples to illustrate the impact of high-energy neutrino research on the particle physics progress of the past years, both in the field of the weak interactions and in that of nucleon structure. How will this develop in the future? I do not know, of course. The increase of proton accelerator energies into the 10 TeV range will certainly permit better QCD tests than those cited above. In general, however, it can be expected that progress in particle physics will depend more and more on colliders, because of their higher centre-of-mass energies. High-energy e-p machines, such as HERA, will permit exploration of inclusive scattering to higher Q^2 domains than will be possible with fixed-target neutrino beams.

However, the fascination with neutrinos and the unanswered questions concerning them- such as their masses-are motivating a broad line of research in astrophysics, accelerator physics, and nuclear physics. One of the first and most important results expected from the two large e^+e^- colliders just coming into operation, the Stanford Linear Collider and the CERN LEP, which will produce lots of Z^0 mesons, is the determination of how many families of leptons and quarks there really are. Are there others besides the three already known? This fundamental question will be answered by determining how often the Z^0 decays to neutrinos, even if the masses of the other members of possible additional families are too large to permit their production at these energies.

REFERENCES

1. F. J. Hasert et al., Phys. Lett. 46B (1973) 138.
2. F. J. Hasert et al., Phys. Lett. 46B (1973) 121.
3. M. Holder et al., Phys. Lett. 71B (1977) 222.
   H. Abramowicz et al., Z. Phys. C28 (1985) 51 and Phys. Rev. Lett. 57 (1986) 298.
4. H. Holder et al., Phys. Lett. **72B** (1977) 254.
5. L. H. Ahrens et al., Phys. Rev. Lett. 54 (1985) 18.
6. J. Dorenbosch et al., Experimental results on electron-neutrino scattering, submitted to Zeitschrift fur Physik C.
7. E. D. Bloom et al., Phys. Rev. Lett. 23 (1969) 930.
   M. L. Breitenbach et al., Phys. Rev. Lett. 23 (1969) 935.
   E. D. Bloom et al., Recent results in inelastic electron scattering, Stanford preprint SLAC-PUB 796 (1970), paper 7b-17 submitted to the 15th Int. Conf. on High-Energy Physics, Kiev, 1970. [See in Proc. talk by R. Wilson (Naukova Dumka, Kiev, 1972), p. 238.1
8. J. D. Bjorken, Phys. Rev. 179 (1969) 1547.
9. R. P. Feynman, Photon-hadron interactions (Benjamin, New York, 1972).
10. J. G. H. De Groot et al., Z. Phys. Cl (1979) 143.
   H. Abramowicz et al., Z. Phys. **Cl2** (1982) 289 and **Cl7** (1983) 283.
   H. Abramowicz et al., A measurement of differential cross-sections and nucleon structure functions in charged-current neutrino interactions on iron, to be submitted to Zeitschrift fur Physik C.
11. A. Benvenuti et al., Phys. Rev. Lett. 34 (1975) 419.
12. M. Holder et al., Phys. Lett. **69B** (1977) 377.
   H. Abramowicz et al., Z. Phys. C15 (1982) 19.
13. S. L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. **D2** (1970) 1285.

Leon M. Lederman

LEON M. LEDERMAN

New York City in the period of 1922 to 1979 provided the streets, schools, entertainment, culture and ethnic diversity for many future scientists. I was born in New York on July 15, 1922 of immigrant parents. My father, Morris, operated a hand laundry and venerated learning. Brother Paul, six years older, was a tinkerer of unusual skill. I started my schooling in 1927 at PS 92 on Broadway and 95th Street and received my Ph.D. in 1951 about one mile north, at Columbia University. In between there were neighborhood junior and senior high schools and the City College of New York. There I majored in chemistry but fell under the influence of such future physicists as Isaac Halpern and my high school friend, Martin J. Klein. I graduated in 1943 and proceeded promptly to spend three years in the U.S. Army where I rose to the rank of 2nd Lieutenant in the Signal Corps. In September of 1946 I entered the Graduate School of Physics at Columbia, chaired by I. I. Rabi.

The Columbia Physics Department was constructing a 385 MeV Synchrocyclotron at their NEVIS Laboratory, located in Irvington-on-the-Hudson, New York. Construction was aided by the Office of Naval Research and "NEVIS" eventually proved to be an extremely productive laboratory, as judged by physics results and students produced.

I joined that project in 1948 and worked with Professor Eugene T. Booth, the director of the cyclotron project. My thesis assignment was to build a Wilson Cloud Chamber. Rabi invited many experts to Columbia to assist the novice staff in what was, for Columbia, a totally new field. Gilberto Bernardini came from Rome and John Tinlot came from Rossi's group at MIT. Somewhat later, Jack Steinberger was recruited from Berkeley. After receiving my Ph.D. in 1951 I was invited to stay on, which I did, for the next 28 years. Much of my early work on pions was carried out with Tinlot and Bernardini.

In 1958, I was promoted to Professor and took my first sabbatical at CERN where I organized a group to do the "g-2" experiment. This CERN program would continue for about 19 years and involve many CERN physicists (Picasso, Farley, Charpak, Sens, Zichichi, etc.). It was also the initiation of several collaborations in CERN research which continued through the mid-70s.

I became Director of the Nevis Labs in 1961 and held this position until 1978. I have been a guest scientist at many labs but did the bulk of my research at Nevis, Brookhaven, CERN and Fermilab. During my academic career at Columbia (1951-1979) I have had 50 Ph.D. students, 14 are professors of physics, one is a university president and the rest with few exceptions, are physicists at national labs, in government or in industry. None, to my knowledge, is in jail. In 1979, I became Director of the Fermi National Accelerator Laboratory where I supervised the construction and utilization of the first superconducting synchrotron, now the highest energy accelerator in the world.

I have three children with my first wife, Florence Gordon. Daughter Rena is an anthropologist, son Jesse is an investment banker and daughter Rachel a lawyer. I now live with my second wife Ellen at the Fermilab Laboratory in Batavia, Illinois, where we keep horses for riding and chickens for eggs. I have been increasingly involved in development via scientific collaboration with Latin America, with science education for gifted children and with public understanding of science. I helped to found and am on the Board of Trustees of the Illinois Mathematics and Science Academy, a three year residence public school for gifted children in the State of Illinois.

Honors: Leon Lederman is the recipient of fellowships from the Ford, Guggenheim, Ernest Kepton Adams and National Science Foundations. He is a founding member of the High Energy Physics Advisory Panel (to AEC, DOE) and the International Committee on Future Accelerators. He has received the National Medal of Science (1965) and the Wolf Prize for Physics (1982) among many other awards.

Honorary D.Sc's have been awarded to Leon M. Lederman by City College of New York, University of Chicago, Illinois Institute of Technology, Northern Illinois University, Lake Forest College and Carnegie Mellon University.

(added in 1991): I retired from Fermilab in 1989 to join the faculty of the University of Chicago as Professor of Physics. In 1989 I was appointed Science Adviser to the Governor of Illinois. I helped to organize a Teachers' Academy for Mathematics and Science, designed to retrain 20,000 teachers in the Chicago Public Schools in the art of teaching science and mathematics. In 1991 I became President of the American Association for the Advancement of Science.

Honors

D.Sc.'s have been awarded among others by the universities at Pisa, Italy and Guanajuarto, Mexico. Elected to the National Academies of Science in Finland and in Argentina. Serves on thirteen (non-paying) Boards of Directors of museums, schools, science organizations and government agencies.

OBSERVATIONS IN PARTICLE PHYSICS FROM TWO NEUTRINOS TO THE STANDARD MODEL

Nobel Lecture, December 8, 1988

bY

LEON M. LEDERMAN

The Fermi National Accelerator Laboratory

I. Introduction

My colleagues Melvin Schwartz and Jack Steinberger and I, sharing the 1988 Nobel Award, were faced with a dilemma. We could, in Rashomon-like fashion, each describe the two-neutrino experiment (as it became known) in his own style, with his own recollections, in the totally objective manner of true scientists. Whereas this could be of some interest to sociologists and anthropologists, this definitely would run the risk of inducing boredom and so we decided on a logical division of effort. Dr. Schwartz, having left the field of physics a decade ago, would concentrate on the origins and on the details of the original experiment. Dr. Steinberger would concentrate on the exploitation of neutrino beams, a field in which he has been an outstanding leader for many years. I volunteered to discuss "the rest," a hasty decision which eventually crystallized into a core theme-how the two-neutrino discovery was a crucial early step in assembling the current world view of particle physics which we call "the Standard Model." Obviously, even a "first step" rests on a pre-existing body of knowledge that could also be addressed. My selection of topics will not only be subjective, but it will also be obsessively personal as befits the awesome occasion of this award ceremony.

I will relate a sequence of experiments which eventually, perhaps even tortuously contributed to the Standard Model, that elegant but still incomplete summary of all subnuclear knowledge. This model describes the 12 basic fermion particles, six quarks and six leptons, arranged in three generations and subject to the forces of nature carried by 12-gauge bosons. My own experimental work brought me to such accelerators as the Nevis Synchrocyclotron (SC); the Cosmotron and Alternate Gradient Synchrotron (AGS) at the Brookhaven National Laboratory (BNL); the Berkeley Bevatron and the Princeton-Penn Synchrotron; the (SC), Proton Synchrotron (PS), and Intersecting Storage Ring (ISR) machines at CERN; the Fermilab 400-GeV accelerator; and the electron-positron collider Cornell Electron Storage Rings (CESR) at Cornell. I can only hint of the tremendous creativity which brought these magnificent scientific tools into being.

One must also have some direct experience with the parallel development of instrumentation. This equally bright record made available to me and my colleagues a remarkable evolution of the ability to record particular subnuclear events with ever finer spatial detail and even finer definition in time. My own experience began with Wilson cloud chambers, paused at photographic nuclear emulsions, exploited the advances of the diffusion cloud chamber, graduated to small arrays of scintillation counters, then spark chambers, lead-glass high-resolution Cerenkov counters, scintillation hodoscopes and eventually the increasingly complex arrays of multiwire proportional chambers, calorimeters, ring imaging counters, and scintillators, all operating into electronic data acquisition systems of exquisite complexity.

Experimentalists are often specialists in reactions initiated by particular particles. I have heard it said that there are some physicists, well along in years, who only observe electron collisions! In reviewing my own bibliography, I can recognize distinct periods, not too different from artists' phases, e.g., Picasso's Blue Period. My earliest work was with pions which exploded into the world of physics (in 1947) at about the time I made my quiet entry. Later, I turned to muons mostly to study their properties and to address questions of their curious similarity to electrons, e.g., in order to answer Richard Feynman's question, "Why does the muon weigh?" or Rabi's parallel reaction, "Who ordered that?" Muons, in the intense beams from the AGS, turned out to be a powerful probe of subnuclear happenings not only in rather classical scattering experiments (one muon in, one muon out), but also in a decidedly non-classical technique (no muons in, two muons out). A brief sojourn with neutral kaons preceded the neutrino program, which my colleagues will have discussed in detail. This led finally to studies of collisions with protons of the highest energy possible, in which leptons are produced. This last phase began in 1968 and was still going on in the 1980's.

Accelerators and detection instruments are essentials in particle research, but there also needs to be some kind of guiding philosophy. My own approach was formed by a specific experience as a graduate student.

My thesis research at Columbia University involved the construction of a Wilson cloud chamber designed to be used with the brand new 400-MeV synchrocyclotron under construction at the Nevis Laboratory about 20 miles north of the Columbia campus in New York City.

I. I. Rabi was the Physics Department Chairman, maestro, teacher of us all. He was intensely interested in the new physics that the highest energy accelerator in the world was producing. At one point I described some curious events observed in the chamber which excited Rabi very much. Realizing that the data was very unconvincing, I tried to explain that we were a long way from a definitive measurement. Rabi's comment, "First comes the observation, then comes the measurement," served to clarify for me the fairly sharp distinction between "observation" and "measurement."

Both experimental approaches are necessary to progress in physics. Observations are experiments which open new fields. Measurements are subsequently needed to advance these. Observations may be qualitative and may require an apparatus which sacrifices detail. Measurement is more usually concerned with the full panoply of relevant instruments. And of course, there are blurred boundaries. In the course of the next 30 or so years I have been concerned with measurements of great precision, e.g., the magnetic moment of the muon', or the mass, charge and lifetime of the muon', measurements of moderate precision like the rho value in muon decay, the elastic scattering of muons', or the lifetimes of the lambda and kaon particles[4]. I have also been involved in observations, which are attempts to see entirely new phenomena. These "observations" have, since 1956, been so labelled in the titles of papers, some of which are listed in chronological order in Table I and as references 5 - 11. I selected these because 1) I loved each one; and 2) they were reasonably important in the evolution of particle physics in the amazing period from the 1950's to the 1980's.

TABLE I. MAJOR OBSERVATIONS

- Observation of Long-Lived Neutral V Particles (1956) Ref. 5.
- Observation of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: The Magnetic Moment of the Free Muon(1957) Ref. 6.
- Observation of the High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos (1962) Ref. 7.
- Observation of Massive Muon Pairs in Hadron Collisions (1970) Ref. 8.
- Observation of π Mesons with Large Transverse Momentum in High-Energy Proton-Proton Collisions (1973) Ref. 9.
- Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions (1977) Ref. 10.
- Observation of the Upsilon 4-Prime at CESR (1980) Ref. 11.

II. Long-Lived Neutral Kaons Observation of a Long-Lived Neutral V Particle'

In 1955, Pais and Gell-Mann" noted that the neutral K meson presented a unique situation in particle physics. In contrast to, e.g., the π^0, the K^0 is not identical to its antiparticle, even though they cannot be distinguished by their decay. Using chargeconjugation invariance, the bizarre particle mixture scheme emerges: K^0 and $\bar{K}^0$ are appropriate descriptions of particle states produced with the well-defined quantum number, strangeness, but two other states, K_L. and K_S, have well-defined decay properties and lifetimes.

The essence of the theoretical point, given in a Columbia University lecture by Abraham Pais in the spring of 1955, was that there should exist, in equal abundance with the already observed K_s (lifetime 10^{-10} sec), a particle with much longer lifetime, forbidden by C-invariance from decaying, as did K_s, into two pions. The clarity of the lecture stimulated what appeared to me to be an equally clear experimental approach, using the cloud chamber which had been invented back in 1896 by the Scottish physicist C.T.R. Wilson. The cloud chamber was first used for making visible the tracks of subatomic particles from nuclear disintegrations in 1911. Supplemented with strong magnetic fields or filled with lead plates, it became the workhorse of cosmic ray and early accelerator research, and was used in many discoveries, e.g., those of the positron, the muon, the lambda, the "0" (now K_s, and K^+. As an instrument, it was more biological than physical, subject to poisons, track distortions, and an interminable period of about one minute. To obtain precise momentum and angle measurements with cloud chambers required luck, old-world craftsmanship, and a large, not-to-be-questioned burden of folklore and recipes. Their slow repetition rate was a particular handicap in accelerator science. Donald Glaser's invention of the bubble chamber and Luis Alvarez's rapid exploitation of it offered a superior instrument for the most purposes and by the mid-50's, very few cloud chambers were still operating at accelerators. At Columbia I had some success with the 11 "-diameter chamber built at the Nevis Synchrocyclotron for my thesis, a comparison of the lifetimes of negative and positive pions[13]. In a stirring finale to this thesis, I had concluded (wrongly as it turned out) that the equality of lifetimes implied that charge conjugation was invariant in weak interactions!

In its history at Nevis, the cloud chamber produced results on the decay of pions[14], on the mass of the neutrino born in pion decay'" (enter the muon neutrino; it would be almost a decade before this number was improved), on the scattering of pions[16], including the first suggestions of

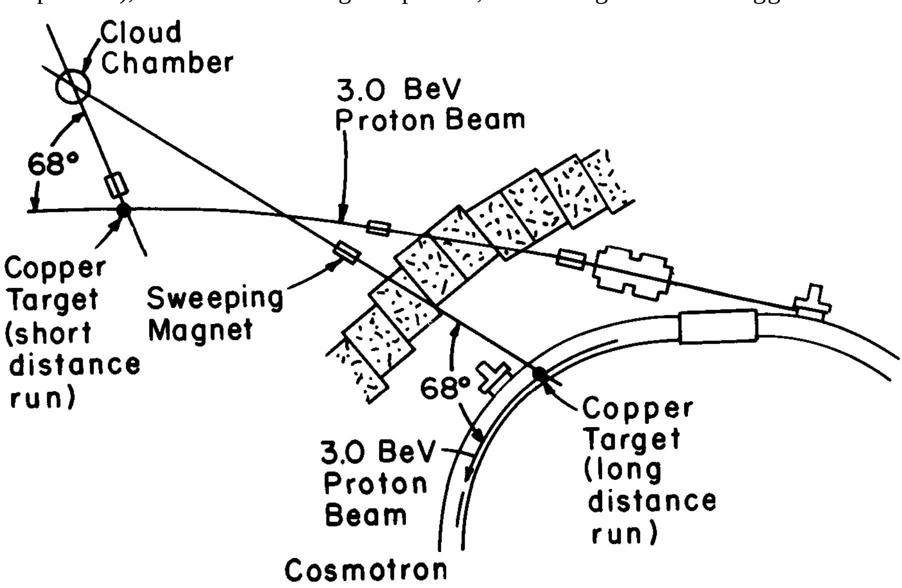


Figure *1.* Experimental arrangements for lifetime study.

strong backward scattering that was later found by E. Fermi to be the indicator of the “3,3” resonance, and on the Coulombnuclear interference of π^+ and π^- scattering in carbon. The carbon scattering led to analysis in terms of complex optical-model parameters which now, over 30 years later, are still a dominating subject in medium-energy physics convocations.

When the Cosmotron began operating in BNL about 1953, we had built a 36”-diameter chamber, equipped with a magnetic field of 10,000 gauss, to study the new Λ^0’s and θ^0’s which were copiously produced by pions of ~ 1 GeV. The chamber seemed ideal to use in a search for long-lived kaons. Figure 1 shows the two arrangements that were eventually used and Figure 2 shows a K_L event in the 36” cloud chamber. The Cosmotron produced

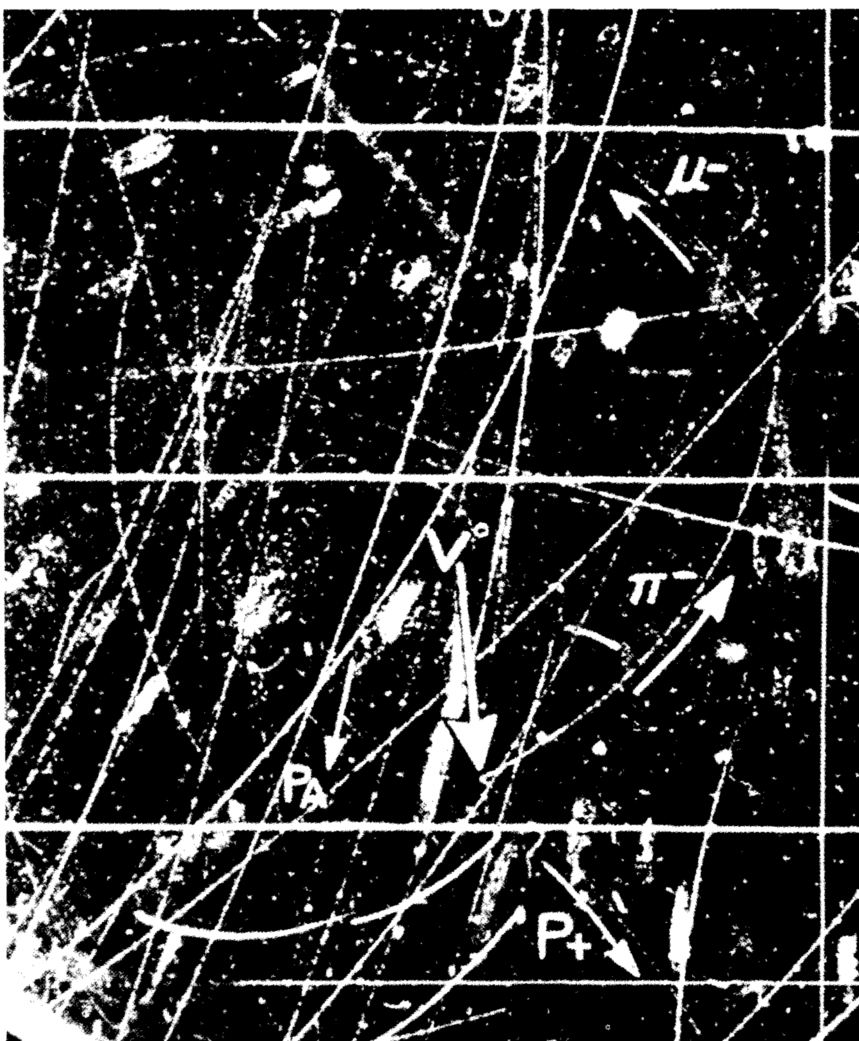


Figure 2. Example of K^0+x+ + π^- neutral particle. *P*+ is shown to be a pion by ionization measurements. P_A is a proton track used in the ionization calibration.

ample quantities of 3-GeV protons and access to targets was particularly convenient because of the magnetic structure of the machine. The trick was to sweep all charged particles away from the chamber and reduce the sensitivity to neutrons by thinning the chamber wall and using helium as chamber gas. By mid-1956, our group of five had established the existence of K_L, and had observed its principal three-body decay modes. Our discussion of alternative interpretations of the "V" events seen in the chamber was exhaustive and definitive. In the next year we measured the lifetime by changing the flight time from target to chamber (both the cloud chamber and the accelerator were immovable). This lifetime, so crudely measured, agrees well with the 1988 handbook value. The K_L was the last discovery made by the now venerable Wilson cloud chamber.

In 1958, we made a careful search of the data for the possibility of a two-body decay mode of K_L. This search was a reflection of the rapid pace of events in the 1956 - 58 period. Whereas C-invariance was the key argument used by Pais and Gell-Mann to generate the neutral K mixture scheme, the events of 1957 (see below) proved that, in fact, C-invariance was strongly violated in weak decays. Since the predictions turned out to be correct, the improved argument, supplied by Lee, Oehme and Yang", replaced C-invariance by CP-invariance, and in fact, also CPT invariance. CP invariance would strictly forbid the decay

$$K_I \rightarrow \pi^+ + \pi^-$$

and, in our 1958 paper based upon 186 K_L, events, we concluded: "... only two events had zero total transverse momentum within errors ... and none of these could be a two-body decay of the K_L^0. An upper limit to $K_L^0 \rightarrow \pi^+ + \pi^-$ was set at 0.6% . . . the absence of the two-pion final state is consistent with the predictions of time reversal invariance."

Six years later, at the much more powerful AGS accelerator, V. Fitch and J. Cronin's, capitalizing on progress in spark chamber detectors, were able to vastly increase the number of observed K_L decays. They found clear evidence for the two-pion decay mode at the level of 0.22 % establishing the fact that CP is, after all, not an absolute symmetry of nature.

The K^0 research eventually provided a major constraint on the Standard Model. On the one hand, it served to refine the properties of the strange quark proposed in 1963 by Gell-Mann. On the other hand, the famous Kobayashi-Maskawa (KM) quark mixing matrix with three generations of quarks was an economical proposal to accommodate the data generated by the K^0 structure and the observation of CP violation. Finally, the neutral K-meson problem (essentially the K_s decay modes) led to the next major observation, that of charge conjugation (C) and parity (P) violation and, together, a major advance in the understanding of the weak interactions. In 1988, neutral K research remains a leading component of the fixed-target measurements at Fermilab, BNL, and CERN.

III. Observation of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays[6]

In the summer of 1956 at BNL, Lee and Yang had discussed the puzzle of the K's (θ, τ puzzle) and were led to propose a number of reactions where possible P violation could be tested in weak interactions[19]. At first glance these all seemed quite difficult experimentally, since one was thinking of relatively small effects. Only C. S. Wu, our Columbia colleague, attempted, with her collaborators at the National Bureau of Standards, the difficult problem of polarizing a radioactive source. When, at a Christmas party in 1956, Wu reported that early results indicated large parity-violating effects in the decay of Co^{60}, it became conceivable that the chain of parity violating reactions: $\pi \rightarrow \mu + \nu$ and then $\mu +\ e + 2\nu$ would not reduce the parity violating effect to unobservability. The "effect" here was the asymmetry in the emission of electrons around the incident, stopped, and spinning polarized muon.

Experience in two key areas set in course a series of events which would convert a Friday Chinese-lunch discussion, just after New Year, 1957, into a Tuesday morning major experimental observation. One was that I knew a lot about the way pion and muon beams were formed at the Nevis cyclotron. In 1950, John Tinlot and I had been pondering how to get pions into the cloud chamber. Until that time, external beams of pions were unknown at the existing cyclotrons such as those at Berkeley, Rochester, and Liverpool. We plotted the trajectories of pions produced by 400-MeV protons hitting a target inside the machine, near the outer limit of orbiting protons, and we discovered fringe field focussing. Negative pions would actually emerge from the accelerator into a well-collimated beam. It remained only to invent a target holder and to modify the thick concrete shield so as to "let them out." In about a month, we had achieved the first external pion beam and had seen more pions in the cloud chamber than had ever been seen anywhere.

The second key area had to do with my student, Marcel Weinrich, who had been studying the lifetime of negative muons in various materials. To prepare his beam we had reviewed the process of pions converting to muons by decay-in-flight. What was more subtle, but easy to play back during the 30-minute Friday evening drive from Columbia to Nevis, was that a correlation of the muon spin relative to its CM momentum would, in fact, be preserved in the kinematics of pion decay-in-flight, resulting in a polarized muon beam. One totally unclear issue was whether the muon would retain its polarization as it slowed from ~ 50 MeV to rest in a solid material. Opportunities to pick up an electron and depolarize seemed very large, but I recalled Rabi's dictum: "A spin is a slippery thing" and decided-why not try it?

Preempting Weinrich's apparatus and enlisting Richard Garwin, an expert on spin precession experiments (as well as on almost everything else), we began the Friday night activities which culminated, Tuesday morning, in a 50 standard deviation parity violating asymmetry in the distribution of

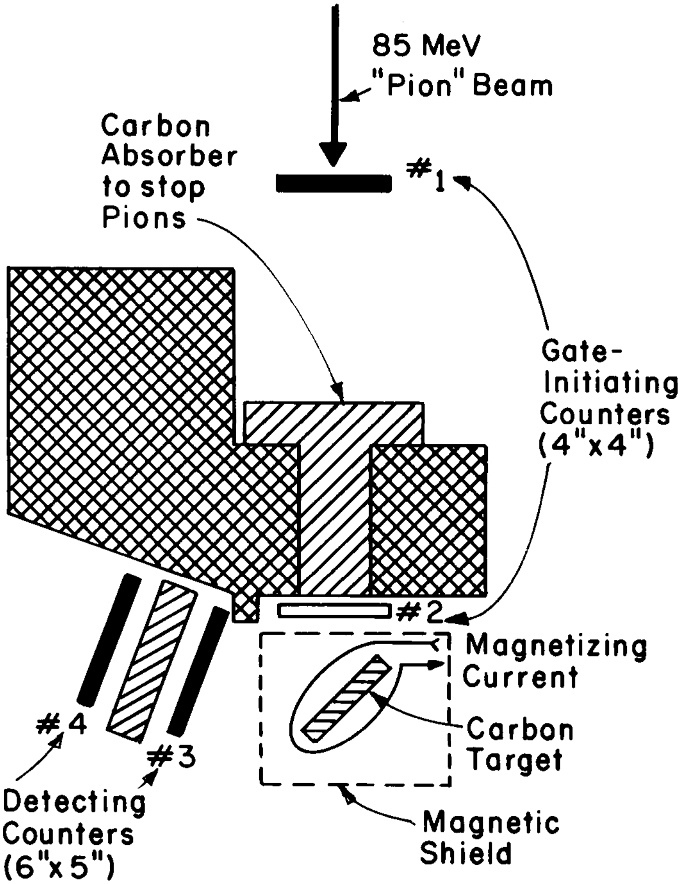


Figure 3. Experimental arrangement. The magnetizing coil was close wound directly on the carbon to provide a uniform vertical field of 79 gauss per ampere.

decay electrons relative to muon spin. Figure 3 shows the very simple arrangement and Fig. 4 shows the data. The following 10 conclusions were contained in the publication of our results:

1. The large asymmetry seen in the CL+ $\rightarrow e^+ + 2v$ decay establishes that the μ^+ beam is strongly polarized.
2. The angular distribution of the electrons is given by

   $1 + a \cos\theta$ where $a = -1/3$ to a precision of 10 %.
3. In reactions $p^+ \mu^+ + v$ *and* $\mu^+ \rightarrow e^+ + 2v$ parity is not conserved.
4. By a theorem of Lee, Oehme, and Yang, the observed asymmetry proves that invariance under charge conjugation is violated.
5. The g-value of the free μ^+ is found to be $+2.00 \pm 0.10$.
6. The measured g-value and the angular distribution in muon decay lead to the strong probability that the spin of the μ^+ is 1/2.
7. The energy dependence of the observed asymmetry is not Strong.
8. Negative muons stopped in carbon show an asymmetry (also peaked backwards) of $a = -1/20$, i.e., about 15 % of that for μ^+.

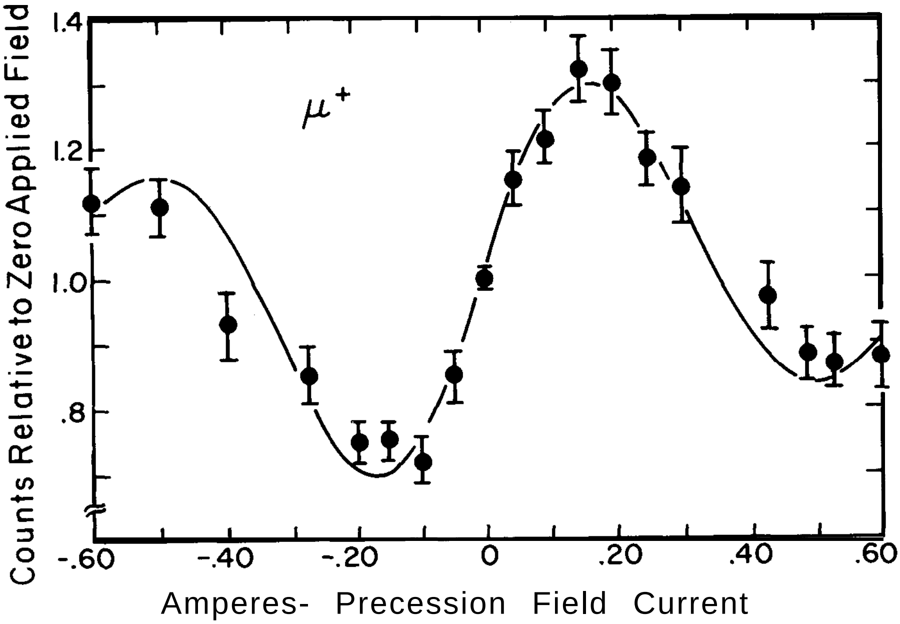


Figure 4. Variation of gated 3-4 counting rate with magnetizing current. The solid curve is computed from an assumed electron angular distribution 1-1/3 cosθ, with counter and gate-width resolution folded in.

9. The magnetic moment of the μ^- bound in carbon is found to be negative and agrees within limited accuracy with that of μ^+.
10. Large asymmetries are found for the e^+ from polarized μ^+ stopped in polyethylene and calcium. Nuclear emulsions yield an asymmetry half that of carbon.

Not bad for a long weekend of work.

This large effect established the two-component neutrinos and this, together with details of the decay parameters as they emerged over the next year, established the V-A structure of the weak interactions. A major crisis emerged from the application of this theory to high energy where the weak cross section threatened to violate unitarity. Theoretical attempts to prevent this catastrophe ran into the absence of evidence for the reaction:

$$\mu \rightarrow e + \gamma$$

The rate calculated by Columbia colleague G. Feinberg[20] was 10^4 times larger than that of the data. This crisis, as perceived by Feinberg, by T. D. Lee, and by Bruno Pontecorvo, provided motivation for the two-neutrino experiment. The stage was also set for increasingly sharp considerations of the intermediate vector boson hypothesis and, indeed, ultimately the electroweak unification.

The 1957 discovery of parity violation in pion and muon decay proved to be a powerful tool for additional research and, indeed, it kept the "pion-factories" at Columbia, Chicago, Liverpool, CERN, and Dubna going for decades, largely pursuing the physics that polarized muons enabled one to do. The earliest application was the precise magnetic resonance measurement of the muon magnetic moment at Nevis in 1957[1]. The high level of precision in such measurements had been unknown to particle physicists who had to learn about precisely measured magnetic fields and spin flipping. A more profound follow-up on this early measurement was the multi-decade obsession at CERN with the g-value of the muon. This measurement provides one of the most exacting tests of Quantum Electrodynamics and is a very strong constraint on the existence of hypothetical particles whose coupling to muons would spoil the current excellent agreement between theory and experiment.

One conclusion of the 1957 parity paper states hopefully that, "... it seems possible that polarized positive and negative muons will become a powerful tool for exploring magnetic fields in nuclei, atoms, and interatomic regions." Today "µSR" (muon spin resonance) has become a widespread tool in solid-state and chemical physics, meriting annual conferences devoted to this technique.

IV. Observation of High-Energy Neutrino Reactions and the Existence of Two Kind of Neutrinos[7]

Since this is the subject of Melvin Schwartz' paper I will not review the details of this research.

The two-neutrino road (a better metaphor would perhaps be; piece of the jigsaw puzzle) to the Standard Model passed through a major milestone with the 1963 quark hypothesis. In its early formulation by both Gell-Mann and George Zweig, three quarks, i.e., a triplet, were believed adequate along the lines of other attempts at constituent explanations (e.g., the Sakata model) of the family groupings of hadrons.

Before the quark hypothesis, a feeling for baryon-lepton symmetry had motivated many theorists, one even opposing the two-neutrino hypothesis before the experiment because "... two types of neutrinos would imply two types of protons." However, after the quark flavor model, Bjorken and Glashow, in 1964[21], transformed the baryon-lepton symmetry idea to quark-lepton symmetry and introduced the name "charm". They predicted the existence of a new family of particles carrying the charm quantum number. This development, and its enlargement by the Glashow, Illiopolis, Maiami (GIM) mechanism in 1970, was another important ingredient in establishing the Standard Model[22].

In GIM, the quark family structure and weak interaction universality explains the absence of strangeness changing neutral weak decays. This is done by assuming a charmed quark counterpart to the second neutrino ν_μ. With the 1974 discovery of the J/ψ at BNL/Stanford Linear Accelerator Center (SLAC) and subsequent experiments establishing the c-quark, the Standard Model, at least with two generations, was experimentally established. Included in this model was the doublet structure of quarks and leptons, e.g.,(u,d),(c,s),(e ν_e),(μ,ν_μ).

The measurements which followed from this observation are given in detail in Jack Steinberger's paper. Major neutrino facilities were established at BNL, CERN, Serpukhov, and Fermilab. Out of these came a rich yield of information on the properties of the weak interaction including neutral as well as charged currents, on the structure functions of quarks and gluons within protons and neutrons, and on the purely leptonic neutrino-electron scattering.

V. Partons and Dynamical Quarks

A. Observation of Dimuons in 30 GeV Proton Collisions"

The two-neutrino experiment moved, in its follow-up phase at BNL, to a much more massive detector and into a far more potent neutrino beam. To provide for this, the AGS proton beam was extracted from the accelerator, not at all an easy thing to do because an extraction efficiency of only 95 % would leave an unacceptably large amount of radiation in the machine.

However, the ability to take pions off at 0^{0} to the beam rather than at the 7^{0} of the original experiment, represented a very significant gain in pions, hence in neutrinos. Thus, the second neutrino experiment, now with healthy competition from CERN, could look forward to thousands of events instead of the original 50.

The major motivation was to find the W particle. The weak interaction theory could predict the cross section for any given mass. The W production was

$$\nu_\mu + A \rightarrow W^+ + \mu^- + A^*$$
$$\bar{\nu}_\mu + A \rightarrow W^- + \mu^+ + A^*.$$

Since W will immediately decay, and often into a charged lepton and neutrino, two opposite-sign leptons appear in the final state at one vertex. Figures 5a, 5b show W candidates. The relatively low energy of the BNL and CERN neutrino beams produced by 30-GeV protons ($\bar{E}$ ~ 1 GeV) made this a relatively insensitive way of searching for W's but both groups were able to set limits

$$M_W > 2 \text{ GeV}.$$

We were then stimulated to try to find W's produced directly with 30-GeV protons, the signature being a high transverse momentum muon emerging from W-decay (~ $M_w/2$). The experiment found no large momentum muons and yielded[23] an improved upper limit for the W mass of about 5 GeV which, however, was burdened by theoretical uncertainties of how W's are produced by protons. The technique led, serendipitously, to the opening of a new field of high-energy probes.

To look for W's, the neutrino-producing target was removed and the beam of protons was transported across the former flight path of 22 m (for pions) and buried in the thick neutrino shield. The massive W could show itself by the appearance of high transverse momentum muons. This beam

Figure 5a. Neutrino event with long muon and possible second μ-meson

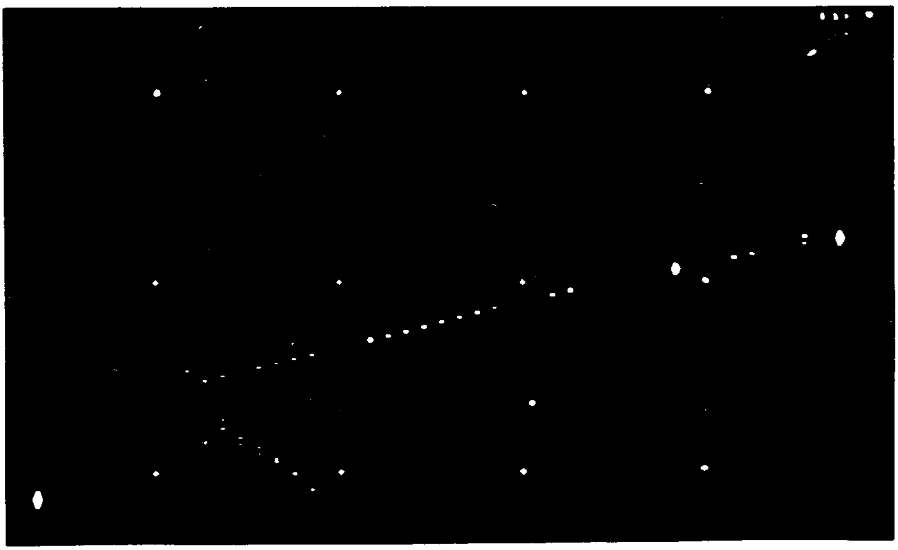


Figure 5b. Neutrino event with long muon and possible electron.

dump approach was recognized in 1964 to be sensitive to short-lived neutrino sources[24], e.g., heavy leptons produced by 30 GeV protons. However, the single muon produced by a hypothetical W could also have been a member of a pair produced by a virtual photon. This criticism, pointed out by Y. Yamaguchi and L. Okun[24], presented us with the idea for a new small-distance probe: virtual photons.

We promptly began designing an experiment to look for the virtual photon decay into muon pairs with the hope that the decreasing yield as a function of effective mass of the observed pair is a measure of small-distance

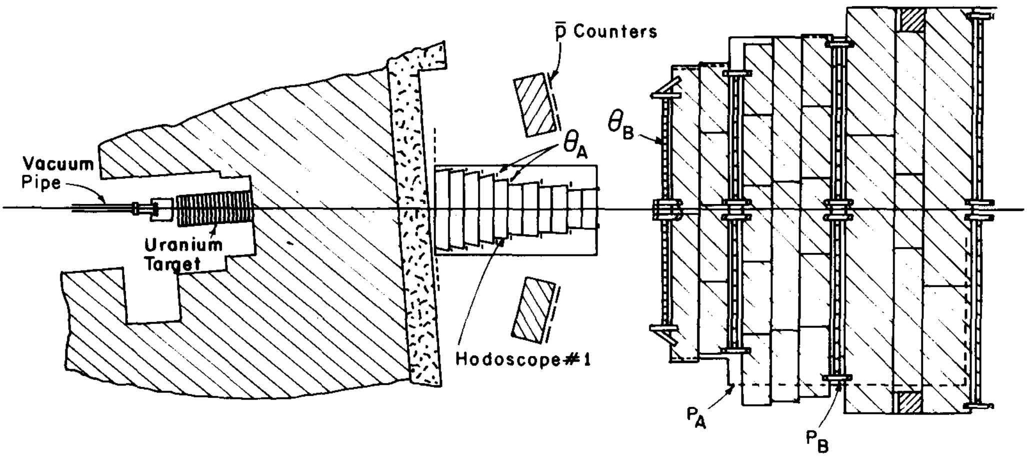


Figure 6. Brookhaven dimuon setup.

physics and that this slope could be interrupted by as yet undiscovered vector mesons. Observation here would be using the illumination of virtual photons whose parameters could be determined from the two-muon final state. In 1967, we organized a relatively simple exploration of the yield of muon pairs from 30-GeV proton collisions. Emilio Zavattini from CERN, Jim Christenson, a graduate of the Fitch-Cronin experiment from Princeton, and Peter Limon, a postdoc from Wisconsin, joined the proposal. Figure 6 shows the apparatus and Figure 7 shows the data. Later we were taught (by Richard Feynman) that this was an *inclusive* experiment:

$$p + U \rightarrow \mu^{+} + \mu^{-} + \text{anything.}$$

The yield of muon pairs decreased rapidly from 1 GeV to the kinematic limit of nearly 6 GeV with the exception of a curious shoulder near 3 GeV. The measurement of muons was by range as determined by liquid and plastic scintillation counters interspersed with steel shielding. Each angular bin (there were 18) had four range bins and for two muons this made a total of only 5000 mass bins into which to sort the data. Multiple scattering in the minimum of 10 feet of steel made finer binning useless. Thus, we could only note that: "Indeed, in the mass region near 3.5 GeV, the observed spectrum may be reproduced by a composite of a resonance and a steeper continuum." This 1968 - 69 experiment was repeated in 1974 by Aubert et al.[25], with a magnetic spectrometer based upon multiwire proportional chambers. The shoulder was refined by the superior resolution into a towering peak (see Fig. 7 a) called the "J" particle.

Our huge flux of 10^{11} protons/pulse made the experiment very sensitive to small yields and, in fact, signals were recorded at the level of 10^{-12} of the total cross section. A crucial development of this class of super-high-rate experiments was a foolproof way of subtracting accidentals.

The second outcome of this research was its interpretation by S. Drell and T-M Yan. They postulated the production of virtual photons by the annihi-

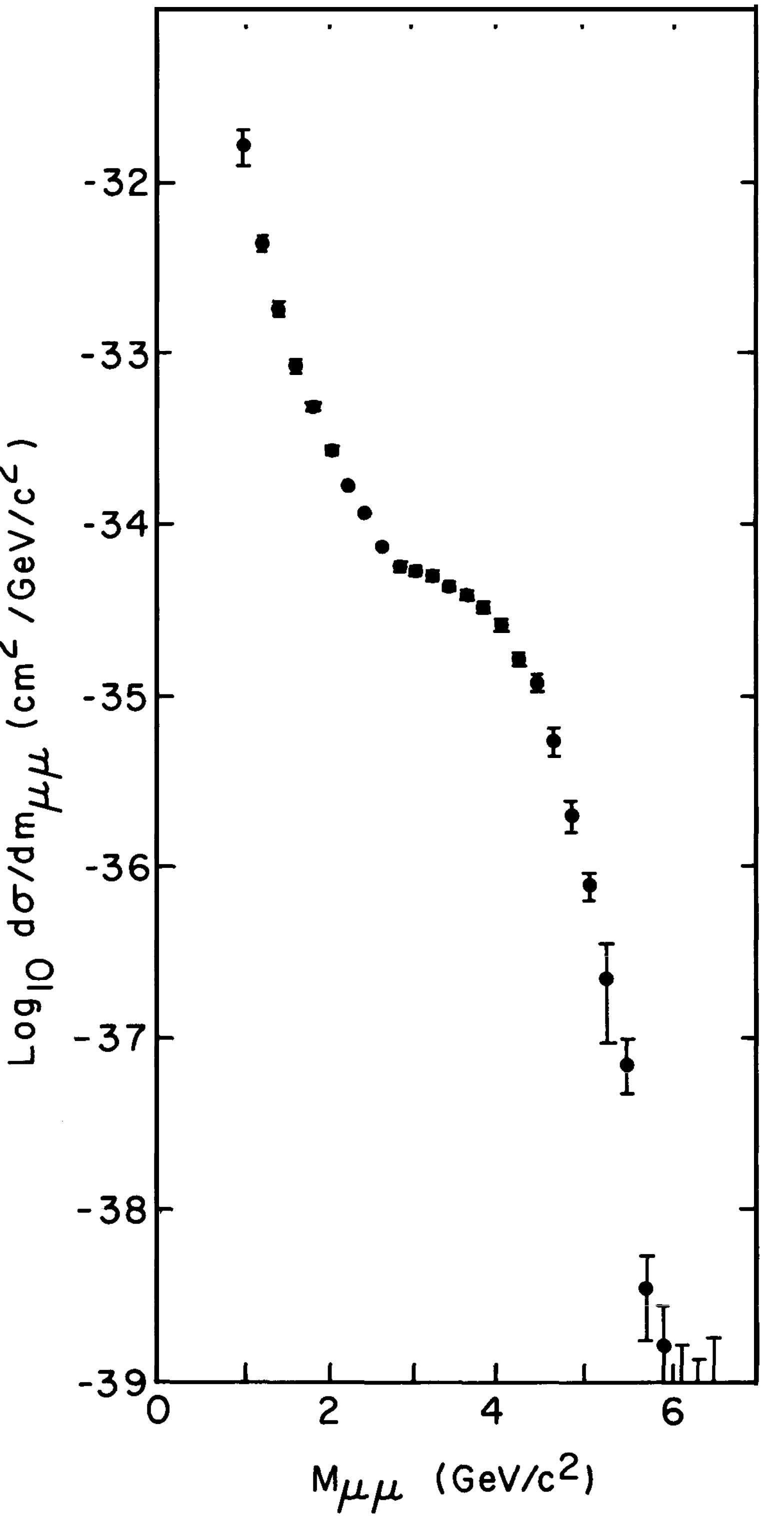


Figure 7a. Data on yield of dimuons vs. mass att 30 GeV.

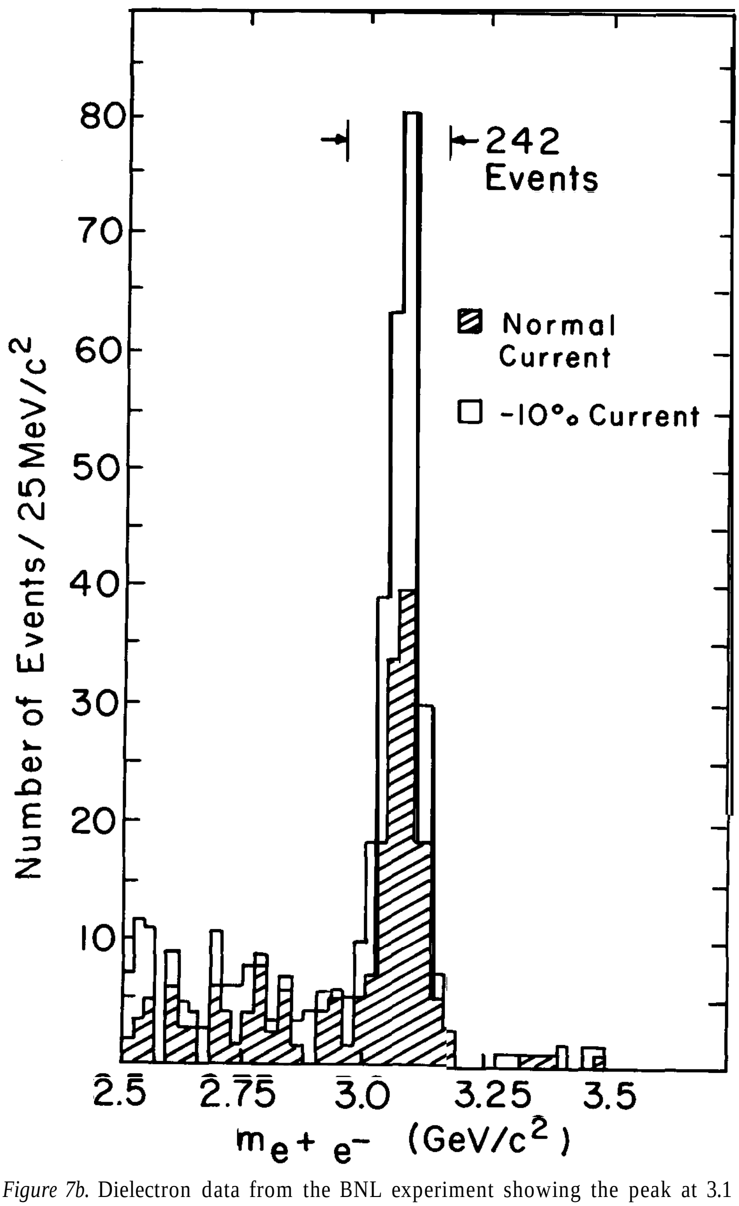


Figure 7b. Dielectron data from the BNL experiment showing the peak at 3.1 GeV which was named "J".

lation of a quark and antiquark in the colliding particles. The application of the now firmly named Drell-Yan process (this is how theorists get all the credit!) in the unraveling of quark dynamics has become increasingly incisive. It lagged behind the deeply inelastic scattering (DIS) analysis by Bjorken and others, in which electrons, muons, and neutrinos were scattered from nucleons with large energy loss. The Drell-Yan process is more dependent upon the strong interaction processes in the initial state and is more subject to the difficult problem of higher-order corrections. However, the dileption kinematics gives direct access to the constituent structure of hadrons with the possibility of experimental control of important parameters of the parton distribution function. Indeed, a very large Drell-Yan industry now flourishes in all the proton accelerators. Drell-Yan processes also allow one to study structure functions of pions, kaons, and antiprotons.

A major consequence of this experimental activity, accompanied by a much greater theoretical flood (our first results stimulated over 100 theoretical papers!), was a parameter-free fit of fairly precise (timelike) data[26] of "two leptons out" to nucleon structure functions determined by probing the nuclear constituent with incident leptons. Some of the most precise data here were collected by the CDHS group of Jack Steinberger and he has covered this in his paper. The agreement of such diverse experiments on the behavior of quark-gluon constituents went a long way toward giving quarks the reality of other elementary particles, despite the confinement restriction.

B. *Observation of π Mesons with Large Transverse Momentum in High-Energy Proton-Proton Collisions*[9]

The dynamics of quark-parton constituents were first convincingly demonstrated by James Bjorken's analysis and interpretation of the DIS experiments at SLAC in 1970. Feynman's parton approach must, of course, also be mentioned. The Berman-Bjorken-Kogut (BBK) paper[27] became the Bible of hard collisionists. In 1971, the brand new ISR at CERN began operations and experimenters were able to observe head-on collisions of 30-GeV protons on 30-GeV protons. The ISR, as the highest-energy machine in the 1970's, was a superb place to practice observation strategy. Impressed by the power of the dilepton proble at BNL and by its hints of structure, Rodney Cool of Rockefeller University and I cooped Luigi DiLella from CERN to help us design an approach which would trade luminosity for resolution. Recall that with the "beam dump" philosophy at BNL we had been able to observe dimuon yields as low as 10^{-12} of the total cross section. However, the penalty was a resolution roughly analogous to using the bottom of a Coca-Cola bottle as the lens for a Nikon. The balance of resolution and luminosity would be a crucial element in the increasing power of the dilepton process.

We learned from Carlo Rubbia about the excellent properties of lead glass as an electromagnetic spectrometer. Photons or electrons would multiply in the high Z medium and dissipate all of their energy in a relatively short length. Improved manufacturing techniques had yielded a dense but transparent glass in which Cerenkov light could be efficiently coupled to good quality photomultiplier tubes. The relatively small response of lead glass to pions and kaons compared to electrons and photons is its great advantage. Six months of hard work in Brookhaven test beams gave us a good command of and respect for this technique and its essential weakness, the calibration process.

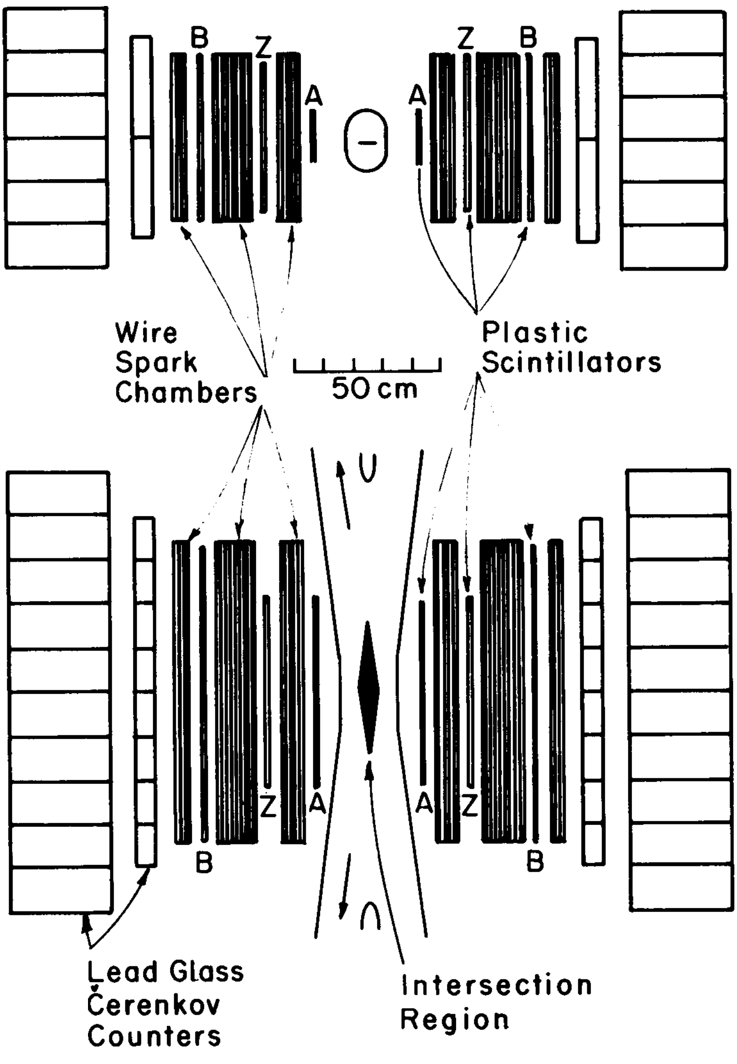


Figure 8. CCR apparatus, CERN ISR.

The idea then was to have two arrays, on opposite sides of the interaction point, each subtending about one steradian of solid angle. Figure 8 shows the CCR apparatus and Figure 9 the data.

The CERN-Columbia-Rockefeller (CCR) team was assembled in 1971 to follow up on the BNI, dilepton results, but now electron pairs where the particles of choice and a large lead-glass array was in place around the interaction point of this very first hadron collider. Here again, the discovery of the J/ψ was frustrated by an interesting background that was totally unexpected but, here again, a new technique for probing small distances was discovered - the emission of high transverse momentum hadrons.

Before the ISR research, a handy rule was that hadron production would fall exponentially with transverse momentum. The CCR result had, at a P_t of

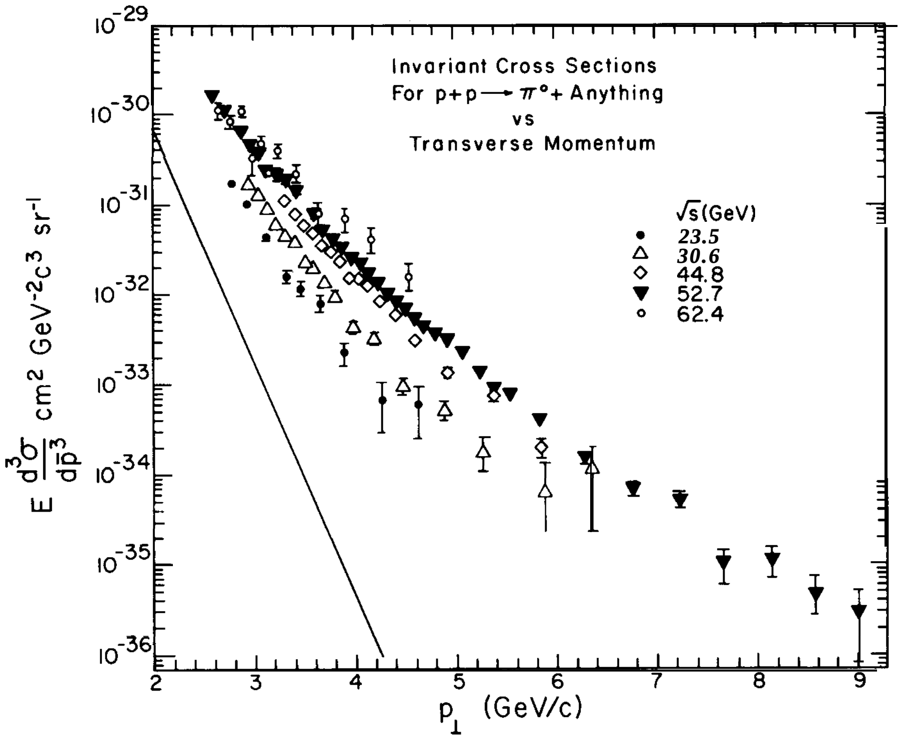


Figure 9. Data from the yield of inclusive π°'s

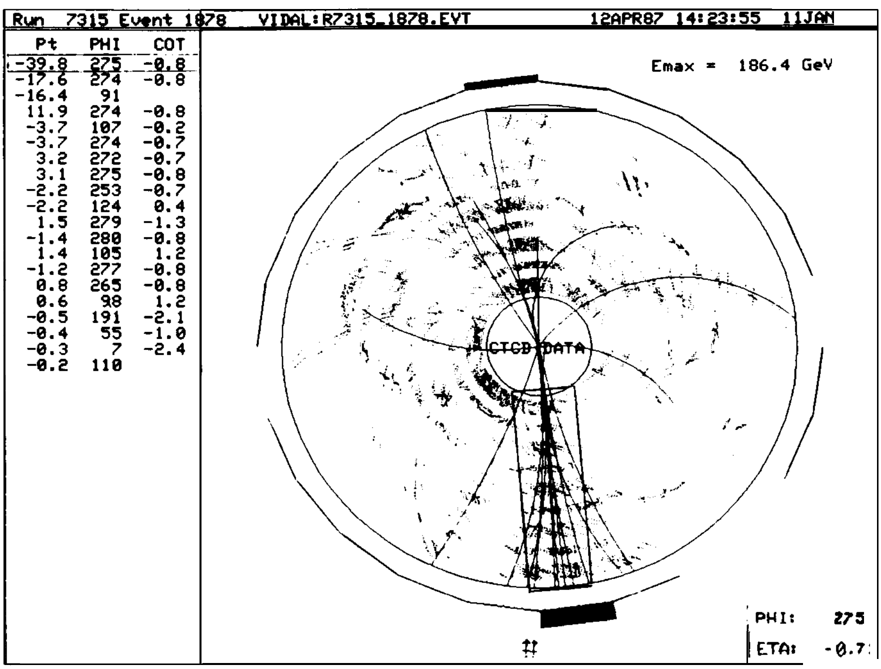


Figure 10. CDF Fermilab dijet at 1.8 TeV.

3 GeV, orders of magnitude higher yield of single π^0's, well detected by the high-resolution lead-glass array. The production rate was observed to be:

$$\sim P_t^{-8} \text{ at } \sqrt{s} = 62 \text{ GeV}$$

which provided a stringent test of the quark-parton model in the early 70's and QCD some few years later. Other ISR experiments quickly confirmed the CCR result, but only CCR had the quality and quantity of data to provide a phenomenological fit. It turned out that one could eventually go directly from these data to parton-parton (or quark-quark, etc.) hard scattering processes. The study of "single inclusive π^0's, at high P_t" evolved into study of the more typical jet structure which now shows up so spectacularly in proton-antiproton collider data. See Figure 10.

Thus, the dilepton adventure using scintillation counters at BNL and the lead-glass exposures to the ISR, initiated independent programs to contribute to the conviction that protons and pions are bound states of confined quarks interacting strongly via the exchange of gluons which are themselves capable of becoming virtual $q\bar{q}$ pairs.

VI. The Third Generation: Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions[10]

In 1969-1970, the BNL dimuon result had not only stimulated the ISR proposal but also a proposal to the Fermilab (then known as NAL and still a large hole-in-the-ground) to do a high-resolution lepton pair experiment. By the time the machine came on in 1972/3, a single-arm lepton detector had been installed, using the very powerful combination of magnetic measurement *and* lead-glass in order to identify electrons with a pion contamination of $\lesssim 10^{-5}$. Such rejection is needed when only one particle is involved.

While the study of "direct" electrons fully occupied the Columbia-Fermilab-Stony Brook collaboration in 1974, the J/ψ was being cheerfully discovered at BNL and SLAC. The single-lepton effects turned out to be relatively unfruitful, and the originally proposed pair experiment got underway in 1975. In a series of runs the number of events with pair masses above 4 GeV gradually increased and eventually grew to a few hundred. During this phase hints of resonant peaks appeared and then disappeared. The group was learning how to do those difficult experiments. In early 1977, the key to a vastly improved dilepton experiment was finally discovered. The senior Ph. D.s on the collaboration, Steve Herb, Walter Innes, Charles Brown, and John Yoh, constituted a rare combination of experience, energy, and insight. A new rearrangement of target, shielding, and detector elements concentrated on muon pairs but with hadronic absorption being carried out in beryllium, actually 30 feet of beryllium. The decreased multiple scattering of the surviving muons reduced the mass resolution to 2%, a respectable improvement over the 10 - 15 % of the 1968 BNL experiment. The filtering of all hadrons permitted over 1000 times as many protons to hit the

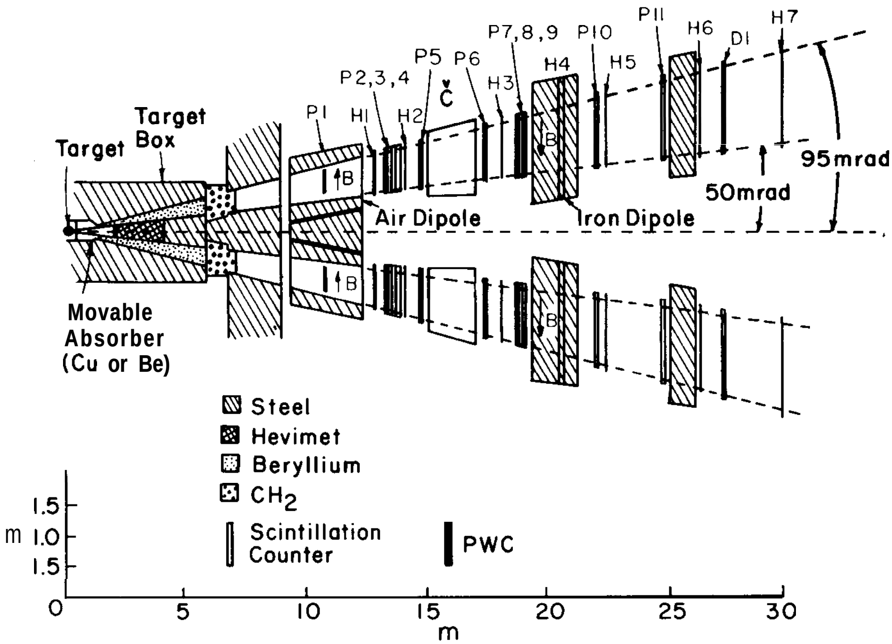


Figure ***11a.*** Plan view of the apparatus. Each spectrometer arm includes 11 PWC's P1-P11, 7 scintillating counter hodoscopes H1-H7, a drift chamber Dl, **and a** gas-filled threshold Ceren-**kov counter C.**

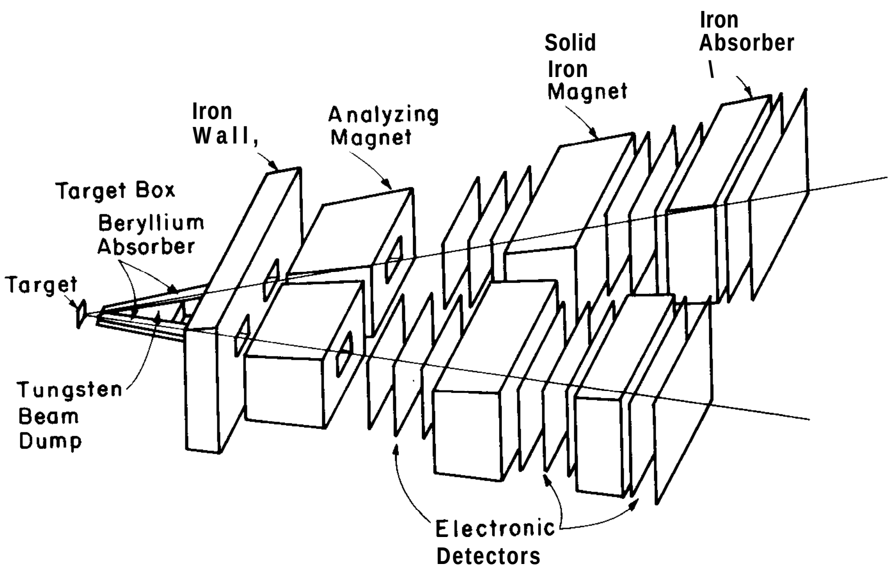


Figure ***11b.*** Schematic sketch of Fermilab dimuon experiment which led to the discovery of the Upsilon particle.

target as compared to open geometry. The compromise between luminosity and resolution was optimized by meticulous attention to the removal of cracks and careful arrangement of the shielding. Recall that this kind of *observation* can call on as many protons as the detector can stand, typically 1 percent of the available protons. The multiwire proportional chambers and triggering scintillators were crowded in towards the target to get maximum acceptance. Muon-ness was certified before and after bending in iron toroids to redetermine the muon momentum and discourage punch-throughs. Figures 11 a, 11 b show the apparatus.

In a month of data taking in the spring of 1977, some 7000 pairs were recorded with masses greater than 4 GeV and a curious, asymmetric, and wide bump appeared to interrupt the Drell-Yan continuum near 9.5 GeV.

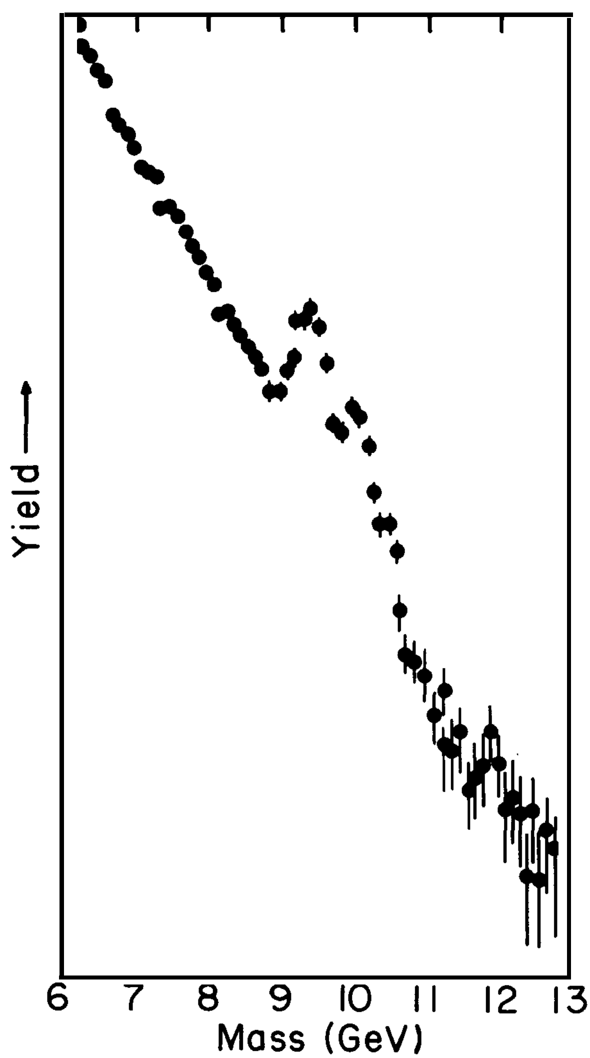


Figure 12a. Peaks on Drell-Yan continuum.

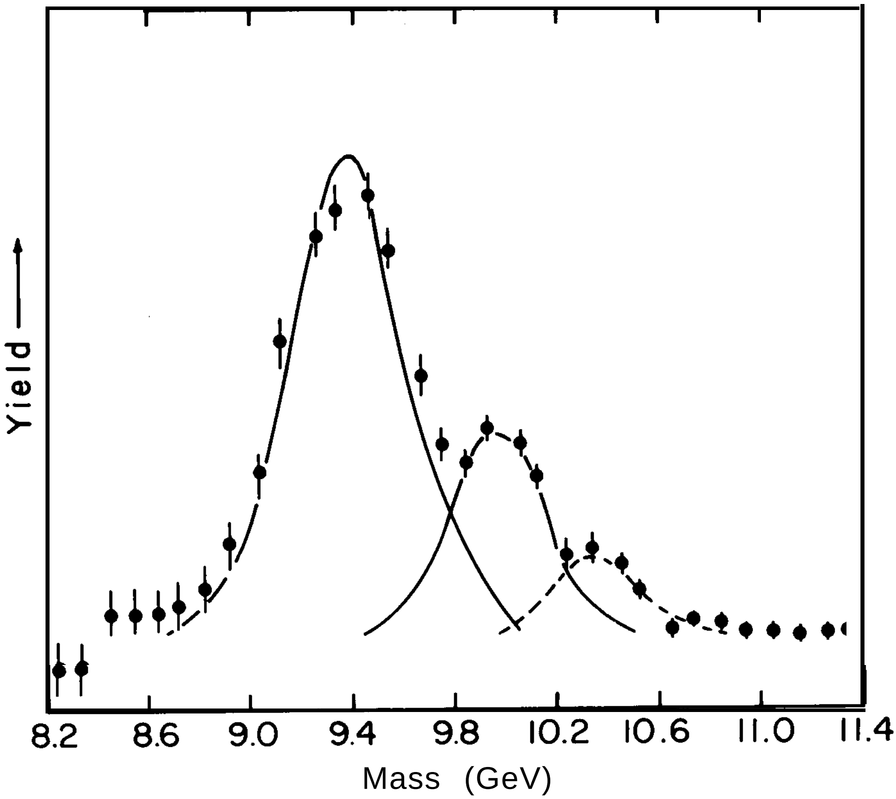


Figure *12b.* Peaks with continuum subracted.

With 800 events in the bump, a very clean Drell-Yan continuum under it and practically no background as measured by looking (simultaneously) for same-sign muons, the resonance was absolutely clear. It was named upsilon and a paper was sent off in August of 1977. By September, with 30,000 events, the enhancement was resolved into three clearly separated peaks, the third "peak" being a well-defined shoulder. See Figures 12a, 12b. These states were called Υ, Υ' and Υ''. Shortly afterwards, the DORIS accelerator in DESY produced the upsilon in e^+e^- collisions and also served to confirm the only plausible interpretation of the upsilon as a bound state of a new quark b with its antiparticle $\bar{b}$. The Υ' and Υ'' were the 2S and 3S states of this non-relativistic "atom". In the Standard Model, we had a choice of charge, + 2/3 (up-like) or - 1/3 (down-like) for the b-quark. The Fermilab data favored - 1/3.

Fallout was relatively swift. Taken together with the discovery by Martin Per1 and his colleague?" of the τ lepton at SLAC slightly earlier, a third generation was added to the Standard Model with the b quark at 5 GeV and the τ -lepton at 2 GeV. This fully confirmed the KM speculation that CP violation may require a third generation. (Clearly we are vastly oversimplifying the theoretical efforts here.)

The bb system was a beautiful addition to $c\bar{c}$ (charmonium) as a measurement laboratory for the study of potential models for the strong quark-quark force. To get in on the fun, I organized a group from Columbia and Stony Brook to design a lead-glass, sodium iodide spectrometer to be used at the CESR machine, ideally suitable for y-spectroscopy. This Columbia, Stony Brook collaboration (CUSB) began taking data in 1979 and soon assisted in the identification of the 4S state[11]. The 4S state is especially important because it is above threshold for hadronic decay to B-states, i.e., mesons having one b quark and a lighter antiquark. Follow-up experiments to learn more about the upsilons were also carried out at Fermilab. These used a number of tricks to even further advance the resolving power without losing luminosity-see Figure 13. By now many other states, including p-states, have been identified in this new heavy-quark spectroscopy.

Recent studies of the B-states in electron-positron colliders indicate that the B system may be far richer in physics than the charm equivalent, the D system. B^0's mix like the K^0 and $\bar{K}^0$ particles. Quoting one of CERN's leading phenomenologists, G. Altarelli: "The observation by Argus at DESY of a relatively large amount of $B^0 - \bar{B}^0$ mixing ... was the most important experimental result of the year [1987] in particle physics." There is the strong possibility that CP violation, seen to date only in the K^0 system, may possibly be observable in the B^0 system. B-factories, usually high-intensity e^+e^- machines, are being proposed in various labs around the world. The Cornell machine is being upgraded to produce of the order of $10^6 B\bar{B}$ pairs a year. Meanwhile the hadron machines are trying hard to solve the very difficult experimental problem of detecting B's (e.g., at the 800-GeV Fermilab fixed target) in a background of 10^6 times as many inelastic collisions. An ambitious detector is being proposed for the Fermilab collider, with the goal of obtaining $10^{10} B\bar{B}$ pairs/year. Judgingfrom 1988 activity, measurements in B-physics will play an increasingly important role in particle research over the next decade. The driving force is the recognition that the third generation seems to be needed to account for CP violation. Taken together with baryon non-conservation, CP violation plays a key role in our understanding of the evolution of the universe, including why we are here. For physicists with a less grandiose view, the quark mixing matrix parameters are part of the basis of our Standard Model and b-physics is the key to these crucial parameters.

The third generation still needs a top quark and as we speak here, searches for this are going on now at the CERN $Sp\bar{p}S$ machine and at the Fermilab collider.

Both machines are operating at very good intensities averaging 200 - 400 nb^{-1} per week. The Fermilab machine has a decided advantage of 1.8 TeV as compared to CERN's 0.63 TeV, but everything depends on the quality of data, the wisdom invested in the design of the detectors and, of course, the mass of the top quark. It does seem safe to predict that a paper will soon appear, perhaps entitled: "Observation of the Top Quark."

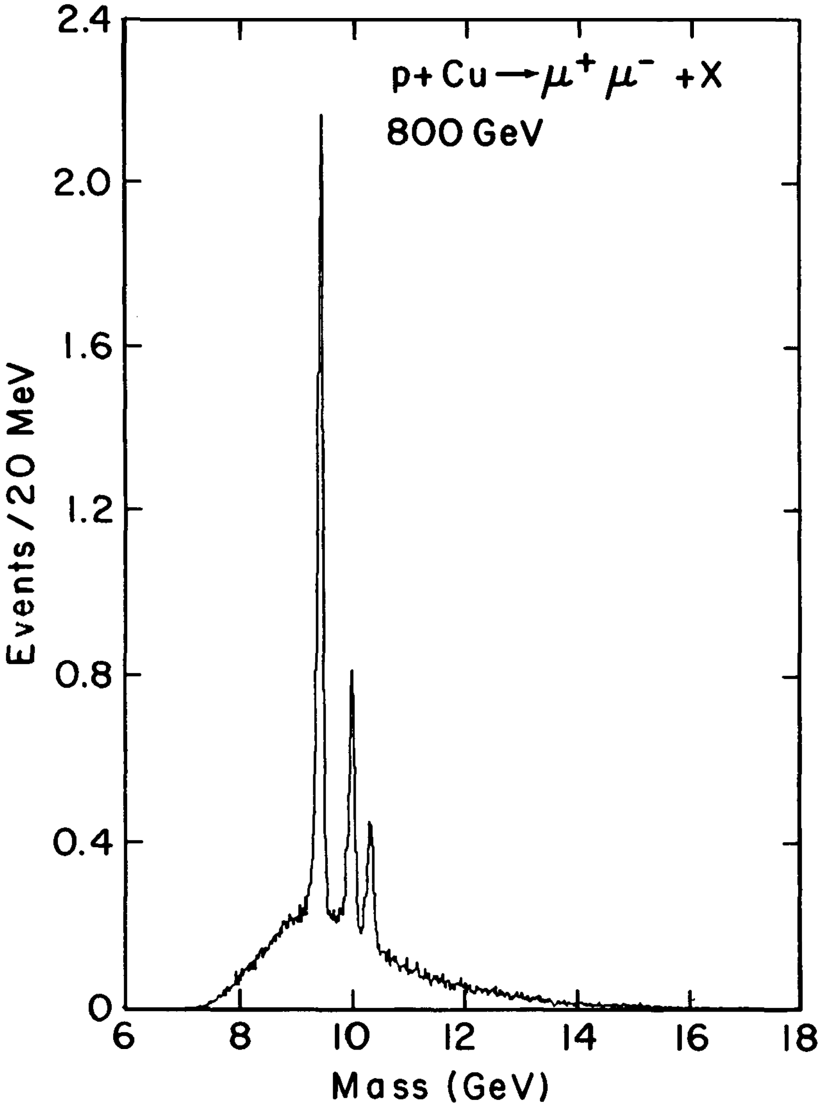


Figure 13. Fermilab E-605 data.

VII. Crucial Issues in Neutrino Physics Today

I conclude this paper with a brief resume of our ignorance about neutrinos. Neutrino interaction data are in good agreement with electroweak theory of the SM and so they will continue to be used to improve our knowledge of quark structure functions, the crucial Weinberg angle, etc. However, we have not yet seen the ν_τ, we do not know if there is a fourth neutrino, we cannot answer urgent questions about the possibility of neutrino mass, and mixing of different flavors, of the stability of the neutrino, whether it has a magnetic moment, and, finally, the nature of the antineutrino, e.g., whether of the Dirac or Majorana type. What makes all of this intensely interesting are two factors: 1) the astrophysical implications of the answers to these questions are awesome; and 2) the view as expressed by Weinberg that ". . . neutrino mass illuminates some of the deepest questions in particle physics.” This is because, in the Standard Model, with the usual quarks, leptons, and gauge bosons, there is no possible renormalizable interaction that can violate lepton number conservation and give the neutrino a mass. Thus, the observation of mass would very likely be a sign of new physics far beyond the Standard Model, perhaps as far as 10^{15} GeV, the scale of Grand Unification.

A. The Third Neutrino, ν_τ

The “three-neutrino” experiment has not been done. Although data from the decay of τ lepton are very strongly suggestive of the existence of ν_τ, direct evidence for ν_τ has yet to appear.

The technical problem is to move the target as close to the detector as possible but to divert the now unstoppable muons by magnetic sweeping. The flux of *v*‘s cannot be predicted with confidence and the shielding configuration is very expensive. This is primarily why the experiment has not yet been done.

B. A Fourth Neutrino?

This question is a shorthand for the issue of the number of generations. Searches for heavier quarks and/or leptons are the *sine qua non* of new accelerators and these have all been negative so far, although the results simply give limits $M_Q > 40$ GeV (same as top quark) and $M_L > 20$ - 40 GeV depending upon the kind of heavy lepton and upon assumptions as to the mass of its accompanying neutrino’“. Important constraints come from astrophysics where the abundance of helium has been related to the number of low-mass neutrinos[29]. Probably one more low-mass neutrino could still be accommodated within the Big Bang nucleosynthesis arguments. The connection between the cosmological model of creation in the Big Bang and the number of generations in the Standard Model is one of the more romantic episodes in the marriage of particle physics and (early universe) cosmology. In fact, one of the strongest supports of Big Bang cosmology is primordial nucleosynthesis; the cooking of the light elements in the caldron beginning at $t \simeq 1$ sec. The astrophysicists manage to get it right: the abundances of deuterium, helium, and lithium. The key is helium 4; its abundance is a sensitive indicator of the total radiation density at formation time. Contributing to this are all the low-mass, relativistic particles, i.e., photons, electrons, and the three neutrinos plus their antiparticles. Another generation containing a low-mass neutrino would probably not destroy the agreement but it would begin to stretch the agreement. Conclusion: there may be a fourth generation, but a fifth generation which included low-mass particles would provide a major problem for our astrophysical colleagues. Of course, there could be something out there which is outside of the generational structure. One experiment soon to yield results is being carried out at the e^+e^- machines at CERN’s Large Electron Positron Collider

(LEP) and the Stanford Linear Collider (SLC) where the width of the Z^0 will give some indication of the number of neutrino pairs into which it can decay. The residual and dominant current interest in the neutrinos comes from astrophysical arguments related to dark matter. This in turn puts the spotlight on the neutrino mass measurements to which we now turn.

C. Neutrino Masses and Oscillation

In the Standard Model, neutrino masses are set to zero and both total lepton number L and lepton flavor number L_i (i = e, μ, τ) are conserved. Neutrino masses "provide a window on the world beyond the SM" and have become one of the outstanding concerns of present-day particle physics. The possibility of oscillation is a statement that $\nu_\mu \rightarrow \nu_e$ is not rigorously forbidden as suggested by our two-neutrino experiment. The issue is given great emphasis by the cosmologists, who are increasingly impinging on the orderly development of particle physics (and what a joy that is!) and by the solar neutrino crisis, which has been around for decades. This is the discrepancy between the number of ν_e's observed to be coming from the sun and the flux that our best knowledge would predict. The detection of ν signals from Supernova 1987a has added to the intensity of interest.

The oscillation possibility was first suggested by B. Pontecorvo in 1967[30]. The neutrino flavor mixing is analogous to the quark mixing as given in the KM matrix. Today, we see many attempts to observe oscillations'. These are at the high-energy accelerator labs, at meson factories, at reactors, and indeed in the solar environment. There, the problem is a theoretical one, to understand the lack of neutrinos from the processes that are known to keep the Sun shining. The solar neutrino crisis alone is receiving the attention of at least 14 large experimental groups around the world and many times that number of theorists!

As of this date, no convincing evidence for oscillations or for neutrino masses has been observed. These indirect evidences for mass differences and other experiments which look directly for neutrino masses are summarized by:

$$m(\nu_e) < \sim 20 \text{ eV}$$
$$m(\nu_\mu) < 0.25 \text{ MeV}$$
$$m(\nu_\tau) < 35 \text{ MeV}$$

Oscillation limits are more conventionally given in terms of limits on the mass differences, A, and the coupled limits on the phase angle, Θ, that defines the mixing strength. Slowly and inexorably the space on the two-dimensional plot (Δ^2 vs sin 28) is being reduced to the lower left hand corner, although logarithmic scales will encourage experimenters to design ever more sensitive tests.

Cosmologists assure us that we live in a universe whose primary component of mass density is dark (non-luminous) and is presently unidentified. Much of this is probably (they say) non-baryonic and some kind of weakly interacting particle carrying some mass (WIMP) is a likely candidate. The principle of minimum complexity would have these be neutrinos and the condition is $\Sigma m_i \sim 20$ eV $(i=e,\mu,\tau)$. This brings the ν_τ forward, as emphasized by Harari who proposes as a matter of urgency a renewed search for $\nu_\mu \rightarrow$ v,.

Other experiments occupying the new pion factories (SIN, TRIUMF, and LAMPF) look for (small) violations of lepton flavor conservation via extremely sensitive searches for such reactions as

$$\mu^+ \not\rightarrow e^+ \gamma \text{ (again but now at B} \sim 10^{-11}\text{!)}$$
$$\text{and } \mu^+ \not\rightarrow e^+ e^+ e^- \text{ (B} < 10^{-12}\text{).}$$

The improvements in experimental techniques and machines conspire to improve these observations by about an order of magnitude every seven years. For completeness we must also list the search for rare decay modes of K-mesons in "kaon factories." Pion, kaon, and B-factories clearly indicate the industrialization of particle physics. The physics objectives of all of these researches are to seek out the tiny influences of presumed new physics which is taking place at the TeV level and higher. For a mature experimenter, these are fun experiments combining the payoff of observations (if and when) with the attention to detail of precise measurements.

To all of the above we should add the new generation structure function research with neutrino beams, probably tagged. Taken together, the 1962 two-neutrino experiment, honored at this meeting, has given rise to a set of activities which, in 1988, continues to play a dominating role in particle physics and its new branches, astrophysics and early universe cosmology.

VIII. *Find Comments*

I would like to conclude this history of the Standard Model, which is not a history at all. From time to time it follows the main road, e.g., when the two-neutrino experiment pointed to flavor and the generational organization of the Standard Model. More often it takes side trails because my own experiments were down those paths. So we have neglected such milestones as the discovery of neutral currents, the τ lepton, the W and Z bosons, charmonium, etc. We have also been crushingly neglectful of the essential theoretical contributions and blitzed through quarks, color, symmetry-breaking, etc.

However, I regret most not having the space to speak more of the accelerators, the detectors, and the people who brought these to be. The Nevis cyclotron was built under the leadership of Eugene Booth and James Rainwater; the AGS, most successful machine ever, led by Ken Green, Ernest Courant, Stanley Livingston, and Hartland Snyder; Fermilab, of course, by Robert Wilson and his outstanding staff. My own detector experience owes much to Georges Charpak of CERN and William Sippach of Columbia. In neglecting these details I am reminded of my teacher, friend, and thesis professor, Gilberto Bernardini, who, when being shown the Nevis cyclotron's innards, exclaimed: "Just show me where the beam comes out." Finally, I make amends to the theorists who are obviously crucial to the entire enterprise. I have enjoyed and profited from many physicists of the theoretical persuasion, but most especially T. D. Lee, M. Veltman, and J. D. Bjorken.

REFERENCES

1. T. Coffin, R. L. Garwin, I.. M. Lederman, S. Penman, A. M. Sachs, Phys. Rev. ***106,*** 1108 (1957).
2. S. Devons, G. Gidal, L. M. Lederman, G. Shapiro, Phys. Rev. Lett. 5, 330 (1960). S. Meyer et al. Phys. Rev. 132, 2963 (1963).
3. R. W. Ellsworth, A. C. Melissinos, J. H. Tinlot, H. von Briesen, Jr., T. Yamanouchi, L. M. Lederman, T. Tannenbaum, R. L., Cool, A. Maschke, Phys. Rev. 165, 1449 (1968).
   C. P. Sargent, M. Rinehart, I,. M. I.ederman, K. C. Rogers, Phys. Rev. 99, 885 (1955).
4. H. Blumenfeld, W. Chinowsky, L. M. Lederman, Nuovo Cimento 8, 296 (1958).
5. K. Lande, E. T. Booth, J. Impeduglia, L. M. Lederman, Phys. Rev. 103, 1901 (1956).
6. R. 1.. Garwin, L. M. Lederman, M. Weinrich, Phys. Rev. 105, 1415 (1957).
7. G. Danby, J-M. Gaillard, K. Goulianos, L. M. Lederman, N. Mistry, M. Schwartz, J. Steinberger, Phys. Rev. Lett. 9, 36 (1962).
8. J. H. Christenson, G. S. Hicks, L. M. Lederman, P. J. Limon, B. G. Pope, E,. Zavattini, Phys. Rev. Lett. 25, 1523 (1970).
9. F. W. Büsser, L. Camilleri, L. DiLella, G. Gladding, A. Placci, B. G. Pope, A. M. Smith, J. K. Yoh, E. Zavattini, B. J. Blumenfeld, L. M. Lederman, R. L. Cool, I,. Litt, S. L. Segler, Phys. Lett. 46B, 471 (1973).
10. S. W. Herb, D. C. Horn, 1,. M. I.ederman, J. C. Sens, H. D. Snyder, J. K. Yoh, J. A. Appel, B. C. Brown, C. N. Brown, W. R. lnnes, K. Ueno. T. Yamanouchi, A. S. Ito, H. Jostlein, D. M. Kaplan, R. D. Kephart, Phys. Rev. Lett. 39, 252 (1977).
11. G. Finocchiaro, G. Giannini, J. Lee-Franzini, R. D. Schamberger Jr., M. Siverta, L,. J. Spencer, P. M. Tuts, T. Bohringer, F. Costantini, J. Dobbins, P. Franzini, K. Han, S. W. Herb, D. M. Kaplan, L. M. Lederman, G. Mageras, D. Peterson, E. Rice, J. K. Yoh, G. Levman, Phys. Lett. 45, 222 (1980).
12. M. Gell-Mann, A. Pais, Phys. Rev. 97, 1387 (1955).
13. L,. Lederman, E. T. Booth, H. Byfield, J. Kessler, Phys. Rev. 83, 685 (1951).
14. L. Lederman, J. Tinlot, E. Booth, Phys. Rev. 81, (1951).
15. L. M. Lederman, Thesis Columbia U, 1951 (unpublished).
16. H. Byfield, J. Kessler, L. M. Lederman, Phys. Rev. 86, 17 (1952).
17. T. D. Lee, Reinhard Oehne, C. N. Yang, Phys. Rev. 106, 340 (1957).
18. J. H. Christenson, J. W. Cronin, V. L. Fitch, R. Turlay, Phys. Rev. Lett. 13, 138 (1964).
19. T. D. Lee, C. N. Yang, Phys. Rev. 104, 254 (1956).
20. G. Feinberg, Phys. Rev. 110, 1482 (1958).
21. J. D. Bjorken, S. L. Glashow, Phys. Lett. ***11, 255*** (1964).
22. S. L. Glashow, J. Iliopoulos, L. Maiani, Phys. Rev. D2, 1285 (1970).
23. R. Burns, G. Danby, E. Heyman, L. M. Lederman, W. Lee, J. Rettberg, and J. Sunderland, Phys. Rev. Lett. 15, 830 (1965).
24. Y. Yamaguchi, Nuovo Cimento ***43A,*** 193 (1966). L. Okun, unpublished (1966).
25. J. J. Aubert, U. Becker, P. J. Biggs, J. Burger, M. Chen, G. Everhart, P. Goldhagen, J. Leong, T. McCorriston, T. G. Rhoades, M. Rohde, Samuel C. C. Ting, Sau Lan Wu, Phys. Rev. Lett. 33, 1404 (1974).
26. D. M. Kaplan, R. J. Fisk, A. S. Ito, H. Jostlein, J. A. Appel, B. C. Brown, C. N. Brown, W. R. Innes, R. D. Kephart, K. Ueno, T. Yamanouchi, S. W. Herb, D. C.

Horn, L. M. Lederman, J. C. Sens, H. D. Snyder, J. K. Yoh, Phys. Rev. Lett. 40, 435 (1978) and A. S. Ito et al., Phys. Rev. D 23, 604 (1981).
27. S. M. Berman, J. D. Bjorken, J. B. Kogut, Phys. Rev. ***D4, 3388*** (1971).
28. Martin Perl: Heavy Leptons in 1986. Proceedings of the XXIII International Conference on High Energy Physics, Berkeley, 1986, p. 596.
29. Gary Steigman, David N. Schramm and James E. Gunn, Phys. Lett. 66B 202 (1977).
30. B. Pontecorvo, Zh. Eksp. Teor. Fiz. 53, 1717 (1967) [Soviet Physics JETP 26,984 (1968)].