PROCEEDINGS
OF THE 1983 JINR-CERN
SCHOOL OF PHYSICS

(Tábor, Czechoslovakia, 5-18 June, 1983)

Volume II

ТРУДЫ
ШКОЛЫ ОИЯИ-ЦЕРН
ПО ФИЗИКЕ 1983

(Табор, Чехословакия, 5-18 июня 1983 года)

Том II

Dubna 1984
CONTENTS

L. Okun

INTRODUCTION TO GAUGE THEORIES. .......................... 3

Preface. .................................................. 3

Lecture 1. Gauge Invariance in Electromagnetic Interaction. ........ 4

Lecture 2. Gauge Invariance in Weak Interaction .................. 11


Lecture 4. Extensions. ...................................... 22

Lecture 5. Teach-Yourself ABC of the Low Energy SUSY. .......... 28

Reprints and Excerpts of Selected Papers
by V. Fock, F. London, O. Klein and H. Weyl. ............ 37

L. Di Lella

HARD HADRON COLLISIONS AT VERY HIGH ENERGY. ............. 82

1. Introduction. ............................................ 82

2. High-p_T Hadronic Final States:
   Historical Survey. ........................................ 82

3. The CERN p\bar{p} Collider Facility and
   Its Two Major Experiments. ............................. 88

4. High-p_T Hadronic Final States at the
   p\bar{p} Collider. ........................................... 96

5. Production and Detection of the W^\pm and
   Z^0 Particles at the CERN p\bar{p} Collider. ............. 112

6. Conclusions. ............................................. 135
INTRODUCTION TO GAUGE THEORIES

L. Okun

Institute of Theoretical and Experimental Physics
Moscow, USSR

Preface

Gauge invariant interactions form the dynamical basis of the modern theory of elementary particles. More than half of 10^4 papers published annually in the field of particle physics deal with gauge invariance. Some parts of the subject are now almost as classic as Euclidian geometry.

In these lectures I will try to be as elementary as possible. I will also try to be complementary to lectures on gauge theories given at previous CERN and CERN-Dubna Schools (B. de Wit, 1982, C. Jarlskog 1981, L. Maiani 1980, ...) and to the lectures given at this School. As a result I will omit some well-known subjects and will give only fragmentary references. (A rather extensive list of references could be found, e.g., in books 1,2). On the other hand, special credit will be given to some old papers, selected reprints of which are presented in the Supplement to these lectures.

Because of lack of space I have skipped in the written text several topics discussed during the lectures and/or in the transparencies:

1. Present and future experiments with W- and Z-bosons.
2. QCD and gluon-gluon collisions.
3. Grand unification and electrical neutrality of atoms.
4. Gauge fantasies on new long range forces.

I take the opportunity to express my gratitude to the hosts of the School for their warm hospitality.
Lecture 1. Gauge Invariance in Electromagnetic Interaction

1.1. Classical Electrodynamics of Classical Particles

Let us begin with the Maxwell equations:
\[ \partial_{\mu} F_{\mu\nu} = j_{\nu}, \]
\[ \partial_{\mu} F_{\mu\nu} = 0, \]
where
\[ F_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu\sigma\tau} F^{\sigma\tau}. \]
The four-potential \( A_{\mu} \) does not appear in these equations explicitly. It enters through the field tensor \( F_{\mu\nu} \):
\[ F_{\mu\nu} = \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu}. \]
The potential \( A_{\mu} \) is determined up to a gradient transformation, which is usually called gauge transformation, namely, \( F_{\mu\nu} \) does not change when
\[ A_{\mu} \rightarrow A_{\mu} + \partial_{\mu} f, \]
where \( f = f(x, y, z, t) \).
Conservation of electric current is necessary for the validity of the Maxwell equations and for gauge invariance. It automatically follows from the definition of \( F_{\mu\nu} \):
\[ \partial_{\mu} j_{\mu} = 0, \quad \text{as} \quad \partial_{\mu} \partial_{\nu} F_{\mu\nu} = 0. \]
The Maxwell equations follow from the Lagrangian
\[ \mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + j_{\mu} A_{\mu}. \]
At first sight this Lagrangian is not gauge invariant. However the extra term, \( j_{\mu} \partial_{\mu} f \), produced by gauge transformation can be easily transformed using the current conservation into a divergence:
\[ j_{\mu} \partial_{\mu} f = \partial_{\mu}(j_{\nu} A^{\nu}), \]
which can be discarded if the function \( f(x, y, z, t) \) has a reasonable behaviour at infinity.

1.2. Electrodynamics of Fields

In field theory the current \( j_{\mu} \) is expressed through charged matter fields (through wave functions of charged particles in quantum mechanics). Consider spin 1/2 and spin 0 charged fields:
\[ j_{\mu} = \frac{e}{c} \bar{\psi} \gamma_{\mu} \psi \]
for Dirac field,
\[ j_\mu = \frac{ie}{
abla^2} \left( \partial_\mu \phi - (\partial_\mu \phi^*) \phi \right) \] for Klein-Gordon field.

(Each that these currents are conserved). The matter and the radiation are described now on the equal footing by the Dirac Lagrangian:

\[ \mathcal{L} = -\frac{1}{4} F_{\mu\nu}^2 + j_\mu A_\mu + i\bar{\psi} \slashed{\partial} \psi - m \bar{\psi} \psi, \]

where \[ \slashed{\partial} \psi = \frac{i}{2} \left[ \gamma \partial_\mu \psi - (\partial_\mu \psi^*) \psi \right]. \]

This Lagrangian can be rewritten by using the covariant (or "long") derivative:

\[ \slashed{\partial} \psi = \partial_\mu - ie A_\mu. \]

\[ \mathcal{L} = -\frac{1}{4} F_{\mu\nu}^2 + i\bar{\psi} \slashed{\partial} \psi - m \bar{\psi} \psi. \]

Also in the scalar case the Lagrangian can be expressed through \( \partial_\mu \):

\[ \mathcal{L} = (\partial_\mu \phi^*) (\partial_\mu \phi) - \frac{1}{2} \Box \phi^* \phi = (\partial_\mu \phi^*) \partial_\mu \phi + \]

\[ + ie [\phi^* \partial_\mu \phi - (\partial_\mu \phi^*) \phi] A_\mu + e^2 \phi^* \phi A_\mu A_\mu - \frac{1}{2} \Box \phi^* \phi. \]

Note the term \( e^2 \phi^* \phi A_\mu A_\mu \), whose presence is dictated by gauge invariance.

The "short" derivative \( \partial_\mu \) and the four-potential \( A_\mu \) enter the Lagrangian only through \( D_\mu \) and \( F_{\mu\nu} \). Note that \( F_{\mu\nu} \) itself can be expressed through a commutator of long derivatives. Consider a product \( \frac{i}{e} \left[ \slashed{\partial}_\mu, \partial_\nu \right] R \), where \( R \) is some (tensor) function of charged fields:

\[ \frac{i}{e} \left[ \slashed{\partial}_\mu, \partial_\nu \right] R = \frac{i}{e} \left[ \partial_\mu (\partial_\nu R) - \partial_\nu (\partial_\mu R) - (\partial_\mu R) \partial_\nu + (\partial_\nu R) \partial_\mu \right] R = \]

\[ = \frac{i}{e} \left[ \partial_\mu (\partial_\nu R) - \partial_\nu (\partial_\mu R) - (\partial_\mu R) \partial_\nu + (\partial_\nu R) \partial_\mu \right] R. \]

In the last expression it is implied that the derivatives act on \( A_\mu \), but not on \( R \). Therefore, we can write in the operator sense:

\[ F_{\mu\nu} = \frac{i}{e} \left[ \partial_\mu, \slashed{\partial}_\nu \right]. \]

In field theory (and in quantum mechanics) the gauge transformation changes not only the potential of the electromagnetic field, but the charged fields as well. By considering the
action of the long derivative one finds that it is invariant if the gauge transformation consists of two parts:

\[ \xi_i \rightarrow \xi_i e^{ieQ_i}, \]
\[ A_\mu \rightarrow A_\mu + \partial_\mu f. \]

1.3. On the Etymology and on the Early History of Gauge Invariance

According to Webster's Dictionary the word gauge is of Old Norman French origin. It has several meanings:

1) a standard measure or scale of measurement,
2) dimensions, capacity, thickness, etc.,
3) any device for measuring something as the thickness of wire, the dimensions of a machined part, the amount of liquid in a container, steam pressure, etc.

Summarizing the content of the preceding sections we can add:

4) gauge in electrodynamics: a phase factor which multiplies the amplitude of a charged matter field (or a wave function of a charged particle) and whose gradient is added to the electromagnetic potential.

In other languages the corresponding words also have all the above meanings: Eich (German), jauge (French), калибр (Russian), калибр (Serbo-Croatian), калибр (Bulgarian), калибр (Czech, Slovenian), kaliber (Polish, Dutch, Danish, Norwegian), calibro (Italian), calibre (Hispanic), mätt (Swedish), merték (Hungarian), chuán (Vietnamese), мээрэн (Mongolic). I apologize for being unable to reproduce Chinese and Japanese hieroglyphs and giving here only their transcription: [cegui] and [gêzi].

Fig. 1 gives an example of a gauge, used to check the diameter of tubes. It is evident that tubes are not invariant with respect to this gauge.

Why are we using the term "gauge invariance" in theoretical physics to describe a symmetry which is definitely not gauge invariance?

The term Eichinvarianz was coined by Weyl in 1919 in the framework of his (unsuccessful) attempt to geometrize the electromagnetic interaction and to construct in this way a unified geometrical theory of gravity and electromagnetism —
"Generalized General Relativity". At that time Weyl used the term Eichinvarianz as a synonym of scale invariance (Massstabinvarianz).

With the advent of Quantum Mechanics, Fock in 1926 has invented the Klein-Fock-Gordon equation (after Klein but before Gordon) and discovered that the equation is invariant with respect to a transformation

\[ A_{\mu} \rightarrow A_{\mu} + \frac{\partial \varphi}{\partial x_{\mu}}, \quad \varphi \rightarrow \varphi e^{i\varphi}. \]

Fock called it gradient transformation.

Note that if \( i \) is dropped, the phase factor becomes a scale factor. This observation was made in 1927 by London, who thus related the phase and gradient transformation to Weyl's old Eichtransformation.

In 1929, Weyl published a paper ("Electron and Gravitation") in which invariance with respect to phase and gradient transformation is stated as a general principle. He called it Eichinvarianz.

It is interesting that all these papers (see Supplement) deal with the construction of unified theory of electromagnetism and gravitation (Fock's paper deals with Kaluza-Klein five-dimensional theory). The really important and everlasting discoveries of these papers (e.g., Weyl spinors) were considered by the authors only as minor by-products on the way to their lofty goal. What will survive from our grand- and super-unification schemes half a century later?
The physical meaning of gauge invariance in Electrodynamics per se does not seem at present to be terribly profound.

For example, a tiny $M_\gamma$ (say $1/M_\gamma \sim 10^{11}$ cm) would destroy the gauge invariance, but all our Earth-bound electrodynamics, including QED, would not be affected.

Consider now the renormalizability of QED. Here gauge invariance is neither necessary nor sufficient: renormalizability would not be destroyed by $M_\gamma \neq 0$, but on the other hand, it would be destroyed by an anomalous magnetic moment term in the Lagrangian, $\mu \leftrightarrow \gamma$ and $\mathcal{L}_\gamma$, despite its explicit gauge invariance (the dimension of $\mu$ being $m^{-1}$).

What is really fundamental in electrodynamics is the conservation of electromagnetic current: conservation of electric charge. Without conservation of electric charge Coulomb's law would be impossible and photon could not be massless (see ref. /1/ and references therein). Unfortunately, the conservation of electric charge is proved experimentally $10^{10}$ times worse than the conservation of baryonic charge:

$$\tau (n \rightarrow p + \text{neutrals}) \gtrsim 10^{22} \text{ years}.$$ New experiments are needed.

In textbooks on Classical Electrodynamics gauge invariance first appeared only in 1941, in the first edition of "Field Theory" by Landau and Lifshitz /3/, which contained a special section: Ch. 16 "Gradient Invariance". But of course the freedom in choosing the form of the potential was exploited long before. In this respect physicists are somewhat like the famous Molière's personage in Le Bourgeois Gentilhomme who suddenly realized that he was using prose in his everyday conservations all his life.

The choice of a potential is like the choice of a coordinate system (see, e.g., ref. /2/). There is a deep analogy between the gauge invariance in Electrodynamics with its freedom to choose the gauge phase locally, and the general coordinate invariance in General Relativity with its freedom to choose locally the coordinate frames. In both cases there is a sort of a "local self-government".
Let us mention here some special gauge conditions widely used in literature:
\[ \partial_\mu A_\mu = 0 \] Lorentz gauge,
\[ \partial_\mu A_\mu = 0 \] Hamilton gauge,
\[ \partial_\mu A_\mu = 0 \] Coulomb gauge,
\[ A_3 = 0 \] axial gauge,
\[ \gamma_\mu A_\mu = 0 \] fixed point gauge.

This last gauge was introduced by Fock /4,5/ and was used as a powerful theoretical tool by Schwinger in his book "Particles, Sources and Fields" /6/. It is easy to check that a potential
\[ A_\mu = \int_0^\Lambda d\lambda \gamma_\mu F_\nu (\lambda x) \]
satisfies the condition \[ \gamma_\mu A_\mu = 0. \]

If a unit time-like four-vector \( J_\mu \) and a unit space-like four-vector \( S_\mu \) are introduced, then the Hamilton, Coulomb and axial conditions can be written in a covariant form:
\[ J A = 0 \] (Hamilton),
\[ A - (J_\omega) J A = 0 \] (Coulomb),
\[ S A = 0 \] (axial).

In many cases gauge invariance can be used to make back-of-the-envelope estimates of cross-sections and rates. Consider for instance, the photon-photon scattering (Fig. 2) in the limit when the frequency of the photon \( \omega \) is small compared with the electron mass \( m \):
\[ \omega \ll m. \]

By virtue of gauge invariance the effective Lagrangian of this process has the form
\[ \mathcal{L}_{\text{eff}} \sim \alpha^2 F_{\mu\nu}^4 m^{-4}, \]
where \( \alpha = 1/137 \) and the factor \( m^{-4} \) is determined by dimensional considerations (\[ [\mathcal{L}] = [m^4], [F_{\mu\nu}] = [m^2]. \]) Now let us take into account the dimension of the cross section:
\[ [\sigma] = [m^{-2}] \] and the fact that the cross section is proportio-
nal to the square of the effective Lagrangian. Then from pure dimensional arguments we get:

\[ \mathcal{L} = a \alpha^4 \omega^6 m^{-8} \]

where \( a \) is a dimensionless coefficient "of the order of unity". Lengthy QED calculations give:

\[ a = \frac{12 \beta \gamma}{3 \gamma} \approx 0.06 \]

Another example is the decay \( \gamma^0 \rightarrow 2\gamma \) (Fig. 3).

The effective Lagrangian in this case is

\[ \mathcal{L}_{\text{eff}} \sim \alpha \left( \frac{\phi}{f} \right) \Gamma \]

where \( \phi \) is the pion wave function and \( f = 130 \text{ MeV} \) is the famous PCAC parameter. Dimensional arguments give us the decay rate in the form:

\[ \Gamma = a \alpha^2 m^3 f^{-2} \]

where again \( a \) is "of the order of unity". Accurate calculation gives:

\[ a = \frac{1}{32\pi^2} \approx 0.07 \]

Now I would like to return to the more general discussion of the gauge invariance. One can often read in textbooks and lecture notes that local gauge invariance of the electron Lagrangian calls for the existence of photon field. Transformation

\[ \psi \rightarrow e^{i\phi} \psi \]

of a free electron Lagrangian produces an extra term

\[ i\varepsilon \left( \partial_\mu \phi \right) \bar{\psi} \gamma^\mu \psi \].

We need the term

\[ i\varepsilon \bar{\psi} \gamma^\mu A_\mu \]

to absorb it.

Some theorists object. One can avoid introducing \( A_\mu \) by giving its role to a derivative of a scalar field \( \phi \) and transforming \( \phi \rightarrow \phi + \bar{f}. \) To be invariant under this transformation, the Lagrangian cannot contain the kinetic term \( (\partial_\mu \phi)^2 \), and the field \( \phi \) enters only through the interaction term \( i\varepsilon \left( \partial_\mu \phi \right) \bar{\psi} \gamma^\mu \psi \). However, because of the vector current conservation this "interaction" is fictitious. So the field \( \phi \) is like a smile of the cat without the cat. And as a matter of fact we have a free electron theory, which is gauge invariant.
Thus, only the nontrivial realization of gauge invariance call for the existence of photons.

I do not know who first made this observation. I heard it from Ogievetsky, Polubarinov, Vainshtein and Khriplovich.

We can arrive at the same conclusion without introducing the field $\psi$. Namely, we can get rid of the term $ie(\partial_\mu \bar{\psi} j^\mu \psi)$ by using the conservation of the current and discarding as a divergence the term $ie \partial_\mu (\bar{\psi} j^\mu \psi)$ in the Lagrangian. Note, that it is precisely this procedure which is used in classical electrodynamics in connection with the transformation $A_\mu \rightarrow A_\mu + \partial_\mu \phi$.

It is proper to conclude this lecture by asking: If gauge invariance is so unimportant, why then is it so important? The answer to this question is: Because photon is not the only gauge particle. Gauge symmetry of electrodynamics is abelian: U(1). Other interactions have non-abelian symmetry groups: SU(2), SU(3) ... They are more restrictive and more profound.

Lecture 2. Gauge Invariance in Weak Interaction

2.1. SU(2) Toy Model

Many important features of the weak interaction gauge theory are reproduced in a simplified model possessing SU(2) gauge symmetry.

Let us begin with an isotopic doublet of two massless fermions $\psi=(\psi_0, \psi_\beta)$ with electric charges $Q_0 = +1/2$, $Q_\beta = -1/2$. The hypercharge (the mean electric charge) of such a doublet is zero, and so we can forget about it and gauge the isotopic group SU(2) only.

The gauge fields $W^+_{\\mu}$, $W^-_{\\mu}$, $W^0_{\\mu}$, form an isotopic triplet, the field $W^0_{\\mu}$ playing the role of the photon that interacts with electric charge $Q = T^3$. Let us define a 2x2 matrix

$$W_\mu \equiv \frac{1}{2} \mathcal{J}^a W^a_\mu \equiv \frac{1}{2} \gamma^\mu W^a_\mu,$$

where $\mathcal{J}^a$ are three Pauli matrices.

The quintessence of the model is the assumption that the Lagrangian is invariant under non-abelian SU(2) gauge transformation:
Comparing this with an abelian electromagnetic case where

\[ S = e^{ie\phi} \equiv e^{ig\phi/\sqrt{2}} \]

we see that a scalar function \( \phi \) is substituted now by three scalar functions \( \phi^1, \phi^2, \phi^3 \), and one U(1) generator \( Q \) - by three SU(2) generators \( \frac{1}{2} \tau^a \).

Now the covariant derivative is a matrix:

\[ \mathcal{D}_\mu = \partial_\mu - ig \, W_\mu. \]

The field strength is also a matrix:

\[ G_{\mu\nu} \equiv \frac{i}{2} \tau^a G_{\mu\nu}^a \equiv \frac{i}{2} \tau^a F^a_{\mu\nu}. \]

In the same operator sense, as in the abelian case, \( G_{\mu\nu} \) is determined by a commutator:

\[ G_{\mu\nu} = \frac{i}{2} [\mathcal{D}_\mu, \mathcal{D}_\nu] = (\partial_\mu W_\nu - \partial_\nu W_\mu - ig [W_\mu, W_\nu]). \]

Note that the last term vanishes in the abelian case but is very important in the non-abelian case. It is trivial to see that under gauge transformation

\[ G_{\mu\nu} \rightarrow G'_{\mu\nu} = SG_{\mu\nu} S^+, \]

The Lagrangian has the form:

\[ \mathcal{L} = -\frac{1}{2} \, T^a \, G_{\mu\nu}^a \, G_{\mu\nu}^a + \bar{\psi} \, \mathcal{D} \, \psi + \bar{\psi} \, \gamma^5 \, \psi. \]

and is gauge invariant.

Note that

\[ -\frac{1}{2} \, T^a \, G_{\mu\nu}^a \, G_{\mu\nu}^a = -\frac{1}{2} \, G_{\mu\nu} \, G_{\mu\nu}. \]

This expression contains not only bilinear terms \( \mathcal{L}^2 \), but trilinear \( \mathcal{L}^3 \) and quadrilinear \( \mathcal{L}^4 \) terms as well (Fig. 4).
The presence of these nonlinear selfcouplings is the salient feature of our toy model. It is also the salient feature of the standard electroweak theory with its SU(2) x U(1) gauge group.

2.2. SU(2) x U(1) Semi-Toy Electroweak Model

Let us now take into account the fact that isotopic doublets in Nature have nonvanishing hypercharge:

\[ Q = T_3 + Y \]

and consider a so to say semi-toy model in which the fermionic doublet consists of two particles: \( \Psi = (\nu, e) \), \( Q_\nu = 0 \), \( Q_e = -1 \), \( Y = -1/2 \).

Now not only isospin but also the hypercharge have to be gauged, and the gauge group is SU(2) x U(1) with four gauge fields ( \( W^+, W^-, W^0 \) and \( B^0 \)) and two coupling constants ( \( g_2 \) and \( g_1 \)):

\[ \mathcal{D}_\mu = \partial_\mu - i g_2 T W^\mu_j - i g_1 Y B^\mu_j. \]

The ratio determines the weak angle \( \theta_w \):

\[ t g \theta_w = g_1 / g_2, \quad \cos \theta_w = g_2 / g, \quad \sin \theta_w = g_1 / g, \]

where \( g = \sqrt{g_1^2 + g_2^2} \).

Two linear superpositions

\[ A = B^0 \cos \theta_w + W^0 \sin \theta_w, \]
\[ Z = -B^0 \sin \theta_w + W^0 \cos \theta_w \]

describe the photon and the Z-boson, respectively. In fact, it is easy to check that

\[ g_1 T_3 W^0 + g_1 Y B^0 = \bar{g} \cos \theta_w \sin \theta_w (T_3 + Y) A + \bar{g} (T_3 \cos^2 \theta_w - Y m^2 \theta_w) Z = \]
\[ = \bar{g} \cos \theta_w \sin \theta_w QA + \bar{g} (T_3 - Q m^2 \theta_w) Z. \]

Hence

\[ e = \bar{g} \cos \theta_w \sin \theta_w QA = g_2 \sin \theta_w, \]

where \( e^2 = 4 \pi L \).

The coupling of the Z-boson to a particle, that has charge \( Q \) and isospin projection \( T_3 \) is
It is important to stress that the formulas are valid when the symmetry $SU(2) \times U(1)$ is unbroken. They are also valid when this symmetry is broken to $U(1)_{em}$.

In the case of unbroken $SU(2) \times U(1)$, when $T_3$ and $Y$ are conserved separately, $W^0$ and $B^0$ are more meaningful than $A$ and $Z$ ($W^0$ being a representation of $SU(2)$, and $B^0$ of $U(1)$). By considering $A$ and $Z$ we make the first step to acknowledging that in Nature $SU(2) \times U(1)$ is broken and only $U(1)_{em}$ survives. The breaking makes $W^\pm$ and $Z$ massive and only the photon stays massless.

2.3. $SU(2) \times U(1)$ Electroweak Interaction

To construct a realistic theory we have to take into account several experimental facts:

(i) we know three generations of fundamental fermions, of quarks and leptons,

(ii) weak interactions are parity-violating; only left-handed helicity states enter the charged weak currents,

(iii) quarks and, at least, some of the leptons are not massless,

(iv) current quarks are superpositions of quark mass eigenstates,

(v) $W$ and $Z$ are heavy.

First of all let us introduce the left-handed ($L$) and right-handed ($R$) fermions:

$$ Y_{L,R} = \frac{1}{\sqrt{2}} \left( 1 \pm \gamma_5 \right) \gamma^\mu. $$

In accordance with (ii), let us assume that $Y^L_{R}$'s are $SU(2)$ singlets, whereas $Y^L_{L}$'s are $SU(2)$ doublets. There are two doublets in each generation:

$$
\begin{align*}
\nu_{eL}, \ & \nu_{\mu L}, \ & \nu_{\tau L}, \\
\epsilon_{L}, \ & \mu_{L}, \ & \tau_{L}.
\end{align*}
$$

$$
\begin{align*}
U_{dL}, \ & d_{L}, \\
U_{cL}, \ & \ell_{L}
\end{align*}
$$

Here $d', s', b'$ are obtained from $d, s, b$ by a unitary transformation with four free parameters: three angles and a phase. Note that because of different values of isospin, the $L$- and $R$-components of the same particle have different values of hypercharge.
We will defer to the next lecture the discussion of the mechanism which gives masses to gauge bosons and fermions. I will only explain here how the masses of W and Z were predicted.

Looking at the diagram (Fig. 5) describing muon decay one easily finds the relation between the effective four-fermion coupling constant, $G^\mu$, the gauge coupling constant, $g_2$, and the mass of the W-boson, $m_W$:

$$\frac{G^\mu}{\sqrt{2}} = \frac{g_2^2}{8m_W^2}.$$ 

The decay rate of the muon is expressed through its mass, $m_\mu$, and $G^\mu$:

$$\Gamma_\mu = \frac{G^\mu m_\mu^5}{192\pi^3}.$$ 

When compared with experimental data on $\Gamma_\mu$ and $m_\mu$, this expression gives the value of

$$G^\mu = (1.66632 \pm 0.00002) \times 10^{-12} \text{ GeV}^{-2}.$$ 

(See page 67 of the Particle Properties Data Booklet, April 1982).

Hence,

$$m_W = \frac{1}{\sin \Theta_W} \left( \frac{E}{12G^\mu} \right)^{1/2} = \frac{37.3}{\sin \Theta_W} G eV.$$ 

The effective four-fermion coupling of neutral currents is caused by the Z-exchange. (See Fig. 5, which describes $\gamma,e^-$ scattering). The corresponding cross section is given by the equation:

$$\sigma_{\gamma,e} = \frac{G^2}{\hat{s}} \hat{s} \left[ \left( -\frac{1}{2} + \sin^2 \Theta_W \right)^2 + \frac{1}{3} \sin^4 \Theta_W \right].$$ 

Here $\hat{s} = (p_e + p_{\nu_e})^2$ and the parameter $s = m_W^2/m_Z^2 \cos^2 \Theta_W$.  

15
In the standard electroweak theory $S=1$; the experimental value cited in the Data Booklet is $\mathcal{S} = 0.992\pm0.020$. The parameter $\sin^2\theta_W$ enters the expressions for the cross sections of other neutral current reactions as well: $\gamma N$-scattering, $eN$-scattering, including the weak interaction of atomic electrons with atomic nuclei, $\mu N$-scattering, and $e^+e^-\rightarrow\mu^+\mu^-$ annihilation. The mean experimental value of $\sin^2\theta_W$ from all these processes, according to the Data Booklet, is

$$\sin^2\theta_W = 0.224\pm0.019.$$ 

This leads to the masses:

$$m_\gamma = 77.9\pm1.7 \text{ GeV}$$
$$m_\mu = 88.8\pm1.4 \text{ GeV}.$$ 

The $W$- and $Z$-bosons were discovered in 1983 by UA1 and UA2 collaborations at CERN. Professors Di Lella and Dydak will provide you with the latest experimental data on these particles.

2.4. On the Early History of Intermediate Bosons

The non-abelian electroweak theory can be traced to several tributaries:

1. Experimental studies of weak interactions.
2. Gauge invariance of electrodynamics.
4. The concept of isospin introduced by Heisenberg in 1932.
5. Yukawa's idea of mesons, the exchange of which gives both the strong forces and the $\beta$-decay (1935).

In 1981 Cecilia Jarlskog /7/ has brought to general attention a paper published by Oscar Klein in 1938. This paper in fact developed an electroweak theory based on the isotopic gauge invariance.

The theory contained two doublets: $p, n$ and $\nu, e$ and three vector particles: $\gamma$, $W^+$ (denoted by $\overline{B}$) and $W^-$ (denoted by $\overline{B}$) with gauge invariant cubic and quartic interactions bet-
ween them. The only coupling constant of the theory was the electric charge $e$. The possibility of neutral currents mediated by the Z-boson (denoted by $C$) was also mentioned.

In building his theory Klein worked in the framework of a five-dimensional Kaluza-Klein world, trying to unite gravity with electromagnetic and nuclear interactions.

Unfortunately, he did not realize that strong and weak interactions are quite different, so he said nothing about the values of the masses of $B$, $\overline{B}$ and $C$-bosons. One is inclined to think that he considered these bosons to play two roles simultaneously: of $W^+$, $W^-$ and $Z^*$ and of $J^+$, $J^-$ and $J^0$, or rather $g^+$, $g^-$ and $g^0$. Furthermore, he was not quite consistent in describing the electromagnetic interaction of nucleons and leptons without introducing the hypercharge. But his equations for the isotopic triplet of gauge fields are absolutely correct.

Klein's theory was firmly forgotten, and the modern non-abelian theories descend from the famous paper by Yang and Mills (1954). Oscar Klein was sixty when this paper was published. In 1953-1965 he served as a member of the Nobel prize committee. He died in 1977. Unfortunately I know nothing about Klein's reaction to the revival of his ideas.

I am grateful to Cecilia Jarlskog for sending me her paper and a xerox copy of Klein's paper, which is reproduced in the Supplement to these lectures.


As we discussed in the preceding lecture, the electroweak gauge symmetry is badly broken in Nature. A fundamental role in this breaking is assigned to hypothetical spinless particles called Higgs bosons or Higgses. It would perhaps be proper to call the Higgs particle higson, like fermion and boson. But "higson" sounds somewhat strange, like "son of Higg". On the other hand, a title in a recent issue of Nuclear Physics "Radiative corrections to Higgs production" looks no less strange. So in my lectures I will refer to the Higgs boson as higgs, with a lower-case h.

3.1. The Minimal Model

Let us begin with a minimal model, containing one isotopic doublet of scalar bosons $\phi = \phi^+, \phi^0$. The part of the Lagran-
gian containing the field $\varphi$ has the form:

$$\mathcal{L}_\varphi = \frac{1}{2} \partial_\mu \varphi \partial^\mu \varphi - \frac{1}{4} \left( |\varphi|^2 - |\lambda|^2 \right)^2 - \frac{\lambda}{2} \left( \bar{\psi}_i \gamma_\mu \gamma_5 \psi_\kappa \right) \varphi + H.c.$$

The term $|\partial_\mu \varphi|^2$ describes free propagation of the field $\varphi$ and its interaction with gauge fields. Here $\partial_\mu = \partial_\mu - i g_2 W_\mu - i g_1 B_\mu$, where $g_2$ and $g_1$ are SU(2) and U(1) gauge coupling constants, $W_\mu = \frac{1}{2} \partial_\mu \varphi^\dagger \varphi$.

The term $-\lambda^2 (|\varphi|^2 - |\lambda|^2)^2$ describes nonlinear selfinteraction of $\varphi$. The coupling constant $\lambda$ is dimensionless and the constant $\frac{\lambda}{2}$ has the dimension of mass.

In the third term $\bar{\psi}_i \gamma_\mu \gamma_5 \psi_\kappa$ are dimensionless Yukawa coupling constants, their indices $i, \kappa$ referring to the flavours of leptons and quarks; H.c. means Hermitian conjugate.

The most important is the sign minus in front of $\lambda^2$: it leads to spontaneous breaking of gauge symmetry. The field $\varphi$ has a nonvanishing vacuum mean value $\langle \varphi \rangle = \frac{\lambda}{2}$. This gives masses to the originally massless $W$- and $Z$-bosons through the term $|\partial_\mu \varphi|^2$ and to leptons and quarks through their Yukawa couplings.

Such spontaneous creation of particle masses guarantees that at high enough momenta, $\sqrt{s} >> \frac{\lambda}{2}$, all the masses can be neglected, and the unbroken SU(2) x U(1) gauge theory reigns at short enough distances.

In the absence of spontaneous breaking we would have four scalar fields: $\varphi = (\varphi^+, \varphi^0), \varphi^0 = (\varphi^-, \varphi^0)$. The electrically neutral components, $\varphi^0$ and $\varphi^0$, are orthogonal superpositions of two real fields

$$\varphi^0 = \frac{1}{\sqrt{2}} (\varphi^1_0 + i \varphi^2_0),$$
$$\varphi^0 = \frac{1}{\sqrt{2}} (\varphi^1_0 - i \varphi^2_0).$$

As a result of breaking, three of the fields

$\varphi^+, \varphi^-, \varphi^0_2$ become longitudinal components of three (now) massive vector bosons $W^+, W^-, Z^0$. There is only one living scalar field left:

$\varphi^0_1 - \frac{\lambda}{2}$.

The field $\varphi$ describes a neutral higgs with the mass $m_H = 2\lambda^2$.

A detailed description of the "lunch", during which the vector fields eat up three of the four scalar fields see, e.g., in ref. [1].
3.2. Properties of Higgs in the Minimal Model

To predict $m_H$, we must know $\lambda$ and $\mu$. It is easy to find $\lambda$. Indeed, $m_H = G_\lambda \sqrt{2}$, so that $\lambda = (2G_\lambda)^{-1/2} \approx 174$ GeV. But the value of $\mu$ is unknown.

If $\lambda$ is small, $\lambda^2 << \mu$, $m_H$ could be close to its lower limit of around 10 GeV. A higgs with a mass of a few tens of GeV's could be produced at SPS $p\bar{p}$-collider and at Tevatron, accompanying the production of W's and Z's, or at LEP in the reaction suggested by Ioffe and Khoze: $e^+e^- \rightarrow ZH$.

The probability of such higgs-bremsstrahlung is of the order of several times $10^{-3}$ for the lowest mass values and rapidly decreases when the mass of higgs increases. The main decay channels of a light higgs are

$$H \rightarrow b\bar{b}, \quad H \rightarrow c\bar{c}, \quad H \rightarrow t\bar{t}.\quad$$

Of high interest is the decay into two gluons, $H \rightarrow gg$, which proceeds through heavy quark loops (Fig. 7).

A light higgs is a "cash register" of heavy quarks, its coupling to two gluons being proportional to the number of quarks $q$ with mass $m_q$, which is large enough: $2m_q > m_H$. There may be not only quarks in the loop, but also other colored heavy particles (quarkinos?).

An inverse process, $gg \rightarrow H$, is an example of what is called the gluon-gluon fusion. The production of a light higgs through gluon-gluon fusion at $p\bar{p}$- or $pp$-colliders has a rather large cross section, $\sigma \sim 10^{-35}$ cm$^2$. Unfortunately, it is difficult to dig out light higgses produced in this way from under a haystack of background events. It is not so for heavy higgses.

If $\lambda^2 > 4\pi \lambda$, then $H$ could be heavier than $W$. Indeed, $m_H \approx m_W$ at $\lambda \approx \frac{1}{4}$. A higgs with $m_H > m_W$ could be produced at LEP II. The cross section of the process $e^+e^- \rightarrow Zh$ (Fig. 8) is expected to be by an order of magnitude smaller than the cross section of the process $e^+e^- \rightarrow W^+W^-$ (Fig. 9). At LEP II
with $\sqrt{s} \approx 200$ GeV one expects $\sigma(e^+e^- \rightarrow ZH) \approx 0.5 \sigma_{\text{point}}$, $\sigma(e^+e^- \rightarrow W^+W^-) \approx 3 \sigma_{\text{point}}$, where $\sigma_{\text{point}}$ is the standard cross section for the electromagnetic process $e^+e^- \rightarrow \gamma \rightarrow \mu^+\mu^-$ in the lowest perturbative approximation (Fig. 10):

$$\sigma_{\text{point}} = \frac{4\pi}{3} \frac{\alpha^2}{s} = \frac{8\pi N_c}{s \cdot (\text{GeV}^2)}$$

where $s$ is the total c.m. energy.

![Fig. 8.](image)

![Fig. 9.](image)

![Fig. 10.](image)

If $\lambda^2 \gg \frac{\mu^2}{s}$ then $m_H \gg m_W$. A heavy higgs, say, with $m_H \gg 2m_W$ is difficult to produce (one needs machines like UNK or SSC) but easy to detect. Its production will often be accompanied by that of Z-, W-boson and it will decay into a pair of Z- or W-bosons; thus its signature is three intermediate bosons. The UA1 and UA2 experiments have revealed that the heavier is a particle the more conspicuous are its high $p_T$ decays. Three intermediate bosons in one collision would be, as Voloshin remarked, a spectacular fireworks.

If $\lambda \gg 1$ and $m_H \gtrsim 1$ TeV we will have a really strong interaction in the Higgs sector and in the sector of W- and Z-bosons (through their longitudinal components).
3.3 Yukawa Couplings in the Minimal Model

The same scalar doublet, which gives masses to \( W \) and \( Z \), can also give masses to fermions. As an example, consider quarks of the first generation: \( u \) and \( d \). Their right-handed components \( u_R \) and \( d_R \), are isotopic singlets; \( u_L \) and \( d_L \) form an isotopic doublet which we denote by \( Q_L \).

\[
Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}.
\]

The Yukawa coupling of the \( d \) quark has the form

\[
f_d \left( \overline{Q}_L d_R \phi + \overline{d}_R Q_L \phi \right),
\]

where

\[
\phi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix}, \quad i = 1, 2
\]

and

\[
\bar{\phi}_i = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix}.
\]

The Yukawa coupling of the \( u \) quark involves a charge-conjugate isotopic doublets, \( \phi_c \):

\[
f_u \left( \overline{Q}_L u_R \phi_c + \overline{u}_R Q_L \phi_c \right),
\]

where

\[
\phi_c = \tilde{\varphi}^\kappa, \quad \overline{\phi}_c = \begin{pmatrix} \varphi^- \\ \varphi^0 \end{pmatrix}.
\]

The use of \( \phi_c \) is necessary because of the conservation of charge and isotopic spin. (In the above expressions tilde (\( \sim \)) denotes charge-conjugate, and bar (\( - \)) denotes hermitian conjugate of a field, the use of antisymmetrical tensor \( \varepsilon^{\kappa \lambda} \) allows one to deal with antiparticles in the same way as with particles). The mean vacuum value of the scalar field, \( \langle \varphi \rangle = \frac{1}{2} \), gives masses to \( u \)- and \( d \)-quarks:

\[
m_u = \frac{f_u}{\frac{1}{2}}, \quad m_d = \frac{f_d}{\frac{1}{2}}.
\]

In the same way the first generation of leptons acquire their masses:

\[
m_e = \frac{f_e}{\frac{1}{2}}, \quad m_\nu = \frac{f_\nu}{\frac{1}{2}},
\]

and so do the fermions of the second and the third generations.

In the framework of three generations not only diagonal mass terms appear, like \( m_u \bar{u} u, m_e \bar{e} e, m_{\mu} \bar{\mu} \mu \), ..., but also non-diagonal ones, like \( m_{u_2} \bar{u} c + m_{u_3} \bar{c} u, m_{e_\mu} \bar{\mu} e \), etc. They originate from inter-generational Yukawa couplings of the type of \( f_{\mu\nu} \bar{\mu} e + f_{\mu\nu} \bar{\nu} e \). It is these non-diagonal masses
that are responsible for the Kobayashi-Maskawa mixing of generations in the charged weak currents: when mass matrix is diagonalized, non-diagonal currents appear. The imaginary parts of the non-diagonal masses are the only origin of CP-violation in the framework of the minimal model.

3.4. What is wrong with the Minimal Model?

We see that scalar fields are responsible for a large number of fundamentally important phenomena. Theoreticians import from Scalarland all the masses, all the mixing angles, and CP-violations. Therefore the discovery of Scalarland by experimentalists should be considered as the task number one.

At the same time the minimal model which we discussed in the last two lectures is obviously very far from being perfect. Too many parameters in it are arbitrary, not fixed by a priori rules. Some of them are fixed a posteriori by experimental data, others are still absolutely free.

In the gauge sector we have two arbitrary gauge couplings: $\alpha_2$ and $\alpha_4$. The parity violation is brought into the model from the outside by postulating that left-handed spinors form isotopic doublets, while the right-handed ones live as singlets. The value of the scalar condensate $\zeta$ is not predicted, the scalar self-coupling $\lambda$ and with it the higgs mass are unknown. There is no principle which determines the pattern of Yukawa couplings. The lack of such a principle is especially painful in the case of neutrino masses (not to mention the $t$-quark mass). We have no reason to believe that neutrinos are massless. On the other hand, we do not understand why they are so light, that is, why their Yukawa couplings are so small.

Only experimentalists can pull Physics out of this valley of sighs.

Lecture 4. Extensions

While awaiting new experimental discoveries theoreticians go ahead building models, which extend the standard minimal model in various directions. These models contain additional gauge bosons, additional fermions, additional higgses. Some of the models have broken global symmetries and Goldstone bosons, i.e., massless spinless particles originating from spontaneous breaking of global symmetries. (I will call these particles goldstons).
In this lecture we will briefly consider some of these extended models.

### 4.1. Extra Gauge Bosons

To introduce extra gauge bosons one has to enlarge the gauge group. The best studied example is the left-right symmetric group (A. Salam, J. Pati, R. Marshak, R. Mohapatra, G. Senjanović and many others): \( SU(2)_L \times SU(2)_R \times U(1) \), which contains, along with our left-handed \( W_L^+ \), \( W_L^- \), and \( Z_L \), three extra bosons \( W_R^+ \), \( W_R^- \), \( Z_R \), which are coupled to right-handed fermion doublets and left-handed fermion singlets. In the unbroken symmetry limit the model is L-R-symmetric, and parity violation appears in it as a result of spontaneous breaking of symmetry. The breaking makes \( W_R \)'s heavier than \( W_L \)'s, so that \( \delta = \frac{(m_{W_L}/m_{W_R})}{\langle \rangle} \). It also mixes \( W_L \)'s and \( W_R \)'s, by a small angle \( \gamma \), thus producing a small coupling of light \( W \)'s with right-handed fermionic currents. We know from \( \beta \)-decay that \( \delta \) and \( \gamma \) are small. According to the standard model the polarization of electrons in allowed \( \beta \)-transitions is equal to \( -\beta = -\nu/c \), where \( v \) is the velocity of the electron and \( c \) is the velocity of light. In the \( SU(2)_L \times SU(2)_R \times U(1) \) model \( \mathcal{P}_F = -\beta(1-2\delta^2) \) if \( \gamma = 0 \), and \( \mathcal{P}_F = -\beta(1-2\delta^2) / (1-\delta^2) \) if \( \delta = 0 \). In the general case, when both \( \gamma \neq 0 \) and \( \delta \neq 0 \), the polarization \( \mathcal{P}_F \) becomes different for the Fermi and the Gamov-Teller transitions, so that the ratio \( \mathcal{R} = \mathcal{P}_F(F)/\mathcal{P}_F(GT) \) is not equal to unity. At present \( \mathcal{P}_F = -\beta \) to within \( \approx 4\% \) for \( F \)-transitions and \( \approx 1\% \) for \( GT \)-transitions (van Klinken et al. /8/). With much better accuracy the right-handed currents could be searched for at the future electron-proton collider HERA. Direct search for the production of heavier gauge bosons is in the programs of Tevatron and UNX.

### 4.2. Extra Fermions

Another class of models, in which parity is violated spontaneously, is that of the so-called vector like models. These models contain, along with our fermions, additional sets of the so-called mirror fermions, which are usually assumed to be much heavier than our fermions. (If we assume that mirror and ordinary fermions have the same masses, then we have to conclude that mirror fermions have their own photons, gluons, and inter-
mediate bosons; see /9/ and references therein). In the vector-like models the "sin" of parity violation could be attributed completely to the fermion (and scalar) sector without involving gauge bosons. Some theorists consider this as a virtue.

4.3. Extra Higgses

There are models in which u- and d-quarks get their masses from two different higgs doublets. In this case there are five physical higgses: $H^+$, $H^-$, $H^0$, $H^{0'}$, $H^{0''}$. The charged higgses should be produced in pairs, electromagnetically:

$$e^+e^- \rightarrow \gamma \rightarrow H^+H^-.$$  

Experimental search for this process at PETRA excludes charged higgses with $m_H < 13$ GeV. (For a recent review see ref. /10/).

Extra higgses should be looked for also among the decay products of $Z$- and $W$-bosons:

$$Z \rightarrow H^+H^-, \quad Z \rightarrow H^0H^{0'}, \quad W^\pm \rightarrow H^0H^C.$$

Of special interest is the reaction

$$e^+e^- \rightarrow Z \rightarrow W^+H^+,$$

where $Z$ is virtual. The point is that the vertex $ZWH$ vanishes in the theories in which all scalar multiplets are doublets. The discovery of this process would signal the existence of scalar multiplet(s) with isospin larger than 1/2.

A neutrino can have a Dirac mass term

$$m\bar{\nu}\nu = m(\bar{\nu}_L \nu_R + \bar{\nu}_R \nu_L),$$

produced by Yukawa term

$$f_\nu(L_L \nu_R \varphi_c + \bar{\nu}_R L_L \varphi_c^L),$$

so that $m = f_\nu \varphi$. (Here $L_L = (\nu_L \nu_L^\dagger)$, and the scalar field $\varphi_c$ has been discussed earlier).

It can have, however, a Majorana mass. For a left-handed neutrino the Majorana mass term, when expressed through Weyl's spinors, has the form

$$m_L (\nu_L^c \nu_L^\beta \varepsilon_{\alpha\beta} + \text{Hermitian conjugate}).$$

Here $\alpha, \beta = 1, 2$ are spinor indices and $\varepsilon_{\alpha\beta}$ is an antisymmetric tensor. The Dirac mass term transforms $\nu_L^\beta$ into $\nu_R^\beta$ (see fig. 11a); the Majorana mass term transforms $\nu_L$ into $\nu_R$ (see fig.
11b) and therefore violates conservation of leptonic charge. Such Majorana mass term can be produced by higgs isotopic triplet $\Phi$ with three components, $\Phi^0$, $\Phi^-$, $\Phi^-$. (Note that one of them is doubly charged). The coupling of this triplet to leptons has the form

$$\mathcal{L}' = \left( \gamma L \nu L \Phi^- + \frac{1}{\sqrt{2}} \left( \nu L \nu L + \nu L e_L \right) \Phi^- + \nu L \nu L \Phi^0 \right) + \text{H.c.}$$

A nonvanishing neutrino mass $m_L = f'_\nu f'_\nu$ is produced by the vacuum mean value $\langle \Phi^0 \rangle = \frac{f'_\nu}{f'_\nu}$. Experimentally $m_L$ of $\nu_e$ is not larger than 30 eV. A new ITEP experiment again gives $m_{\nu_e} \approx 30 \text{ eV} \pm \text{ a few eV}$, but up to now there were no independent measurements of comparable accuracy which could confirm or disprove this value. With $\frac{f'_\nu}{f'_\nu} \sim 10^2$ GeV this means that $f'_\nu, f'_\nu \sim 10^{-10}$. There is no explanation why they should be so small.

Another possibility, which is actively discussed, is that $\nu_L$ is massless ($m_L \to 0$) but $\nu_R$ is very massive ($m_R \geq 10^{11}$ GeV). $\nu_L$ and $\nu_R$ are mixed by the Dirac mass $m \gtrsim 1 \text{ MeV}$, the mass matrix being

$$\begin{pmatrix}
0 & m \\
 m & m_R
\end{pmatrix}.$$ Then the mass of the lightest mass-eigenstate is $m \approx m^2 / m_R$.

Majorana neutrinos trigger the neutrinoless double $\beta$-decay (see fig. 12) and an intricate pattern of neutrino oscillations.

![Diagram](attachment:fig11.png)

Fig. 11.

![Diagram](attachment:fig12.png)

Fig. 12.

Both phenomena are eagerly searched for, but without any positive results. But let us end this digression on neutrino masses and return to scalar bosons.
With several scalar multiplets it is possible to construct models in which CP is violated not (only) through the Yukawa couplings, but (also) in the higgs sector. The mechanism of this violation could be either explicit (complex coupling constants of nonlinear interactions between scalars) or spontaneous. In the latter case the Lagrangian is CP-invariant, all coupling constants being real, but the condensates - the vacuum mean values of scalar fields - are complex and thus CP-noninvariant.

If the only source of CP-violation is in the higgs sector, then it could be shown (see ref. /11,12/) that the dipole moment of the neutron, $\Delta n$, has to be of the order $10^{-25}$ cm, which is on the brink of contradiction with experimental data. Moreover in this case charged higgses have to be rather light ($m_H \lesssim 10$ GeV), which, as we have also seen, does not seem to be borne out by experiments.

Let us mention here that if CP is violated spontaneously and the vacuum is characterized by a phase factor $e^{i\delta}$, it is always possible to have a complex conjugate vacuum with $e^{-i\delta}$. That means that vacuum domains with alternative signs of CP-violating phases have to be considered. These domains would be separated by very thin but very massive walls, which would influence cosmological history of the universe.

### 4.4. Goldstones and Pseudogoldstons

The spinless products of the breaking of a local symmetry are called higgses. Higgses are massive. The spinless products of the breaking of a global symmetry are called goldstons. Goldstons are massless. If the spontaneously broken symmetry is from the beginning an approximate one (broken explicitly or by quantum anomalies) then a goldston becomes a pseudogoldston. Pseudogoldstons have nonvanishing masses.

An example of composite (non-elementary) pseudogoldstons are pions. Pions are pseudogoldstons produced by the dynamical breaking of global chiral isotopic invariance of QCD. This invariance is an approximate one, since the u- and d-quarks are light but not massless. Were u- and d-quarks massless, pions also would be massless.

Except for pions and their SU(3)-relatives $K^-$ and $\bar{\Lambda}$-mesons...
no other (pseudo)goldstons have been observed experimentally. Up to now they live only on the pages of theoretical papers. I will mention here some of these hypothetical particles: axion, arion, majoron, familon, technipion... Axion /13-18/ corresponds to the global chiral U(1)-symmetry in the quark sector. Arion /19-20/ corresponds to the global chiral U(1)-symmetry in the lepton sector. Familons /21,22,23/ correspond to the so-called horizontal symmetry, relating fermions of different generations. Majoron /24,25/ corresponds to the global conservation of the leptonic charge. All these particles are assumed to be produced by spontaneous breaking of the symmetries involved. Technipions /26,27/ are pseudogoldstons that are produced by the dynamical breaking of global isotopic chiral symmetry of a hypothetical technicolor interaction. The interaction of a goldston $\Phi$ with the relevant current $j_{\mu}$ (the conservation of which is broken) is described by an expression

$$\frac{1}{\sqrt{\beta}} \left( \partial_{\mu} \Phi \right) j_{\mu}.$$ 

The constant $\sqrt{\beta}$ has the dimension of mass and characterizes the scale at which the symmetry is broken.

From the experimental absence of the decay

$K^0 \rightarrow \pi^+ + \text{familon}$

one can deduce /21-23/ that $\sqrt{\beta} > 10^{10}$ GeV for the global horizontal symmetry between d, s, b quarks. Somewhat lower limit for $\sqrt{\beta}$ could be obtained from the absence of the decay

$\mu \rightarrow e + \text{familon}$

in the case of the horizontal $e-\mu - \tau$-symmetry.

The exchange of massless goldstons would lead to long-range forces. These forces would be especially interesting in the case of diagonal vertices that transform a fermion into itself. Such diagonal vertices would generate low-energy long-range interactions between particles of stable matter. It is easy to see that the vector part of the diagonal vertex, $\gamma \gamma \gamma \gamma$, would give a vanishing contribution because of the equation $\partial_{\mu} \gamma \gamma \gamma \gamma = 0$. The nonvanishing contribution would be given by the axial-vector part, $\gamma \gamma \gamma \gamma$. In the static limit

$$\frac{1}{\sqrt{\beta}} q q q \gamma \gamma \gamma \gamma = \frac{1}{\sqrt{\beta}} (q q q) (\bar{q} \bar{s} \gamma),$$

27
where \( q \) is the four-momentum of the goldston (arion) and \( \vec{q} \) is its spatial component. The static potential due to virtual arion exchange is like the potential between two magnetic moments due to virtual photon exchange \(^{19,20} \): 
\[
\frac{(\vec{q} \cdot \vec{\sigma})(\vec{q} \cdot \vec{\sigma})}{V^2 q^2} \rightarrow \frac{\vec{s} \cdot \vec{\sigma}}{V^2} \left( \frac{\vec{s} \cdot \vec{\sigma}}{V^2} - \frac{3 \vec{s} \cdot \vec{\sigma}}{V^2} \right).
\]

In a recent experiment \(^{28} \) the ratio of nuclear spin precession frequencies of two mercury isotopes were measured. The change of this ratio with switching on of an additional magnetic field, screened by a ferromagnetic screen was searched for. If arions exist the magnetic field by acting on the screen would produce an arionic magnetic field. The idea of the experiment was based on the observation that arionic magnetic moments of the nuclei are not proportional to their magnetic moments. The negative result of this experiment gives:

\[
V_e V_q > 10^8 \text{ GeV}^2.
\]
(Here index e stands for electron and q for u- and d-quarks).

This negative result was not an unexpected one, because astrophysical limits are even more stringent. By considering the energy flux, which inner regions of the Sun and red giants would radiate in the form of arions, one can conclude that

\[
V_e > 10^6 \text{ GeV (Sun)}, \quad V_q > 10^9 \text{ GeV (red giants)}.
\]

Nonobservation of axions and arions may mean that there are no spontaneously broken symmetries or that the breaking occurs at very short distances (say, \( \sqrt{\ } \approx 10^{10} - 10^{11} \text{GeV} \)).

Lecture 5. Teach-Yourself ABC of the Low Energy SUSY

5.1. Hierarchy Problem

One of the most fundamental problems, mentioned in the previous lecture, is the problem of mass scale, which is referred to as hierarchy problem. What determines the value of the W-boson mass? The phenomenological answer is: Fermi coupling constant \( G_F \). But what determines the value of this constant in the theory? The answer is: the value of the vacuum mean value \( \langle \sigma \rangle \) of the scalar field. But what determines the value of \( \langle \sigma \rangle \), what determines the mass scale of the scalar sector?
If we consider the contribution to the mass of the scalar field given by Fig. 13, we find that it is divergent. In a renormalizable theory there is no natural cut-off value for this divergence and the first place where it can be cut (from the outside) is the so-called Planck mass, \( m_{\text{Pl}} \), where gravitational interaction becomes strong. (As you perhaps know, \( m_{\text{Pl}} \sim G_N^{-1/2} \approx 10^{19} \text{ GeV} \), where \( G_N \) is the famous Newton constant). But in this case we would get \( \frac{1}{\lambda} \sim m_{\text{Pl}} \) and \( G_F \sim G_N \), while experimentally \( G_N/G_F \sim 10^{34} \). This extremely small ratio, this hierarchy of scales, is a real challenge to theorists.

A possible way to avoid the above non-stop flight to the Planck mass is an accurate compensation between the bosonic loop of Fig. 13 and the fermionic loop of Fig. 14. To achieve such compensation one needs a sort of symmetry between bosons and fermions. Another physical quantity which calls for boson-fermion compensation is the so-called cosmological term - the energy-momentum tensor of the vacuum. In this case the compensation has to be miraculously accurate.

The simplest form of boson-fermion symmetry is the so-called \( N = 1 \) supersymmetry (\( N = 1 \) SUSY), in which there is one superpartner for each known particle.

5.2. Superparticles

We will use for a superpartner of a known particle suffix \( \text{ino} \) and decorate the corresponding letter by a (super)hat or (super) tilde. (The hat may be more convenient than tilde, as the latter is often used in the literature to denote antiparticles).
Let us begin with spin-zero "inos": leptinos and quarkinos, which are often referred to in the literature as sleptons and squarks:

\[
\begin{align*}
e &\rightarrow \hat{e} \quad \text{(eino, or electrisno, or selectron)} \\
\not{\nu} &\rightarrow \hat{\nu} \quad \text{(nuino, smu, smuon)} \\
\not{\tau} &\rightarrow \hat{\tau} \quad \text{(tauino, stau)} \\
\not{\mu} &\rightarrow \hat{\mu} \quad \text{(e-muino, e-snu)} \\
\not{\nu} &\rightarrow \hat{\nu} \quad \text{(mu-nuino, mu-snu)} \\
\not{\tau} &\rightarrow \hat{\tau} \quad \text{(tau-nuino, tau-snu)} \\
u &\rightarrow \hat{\nu} \quad \text{(uiino, u-squark)} \\
d &\rightarrow \hat{d} \quad \text{(dino, d-squark)} \\
s &\rightarrow \hat{s} \quad \text{(sino, s-squark)} \\
c &\rightarrow \hat{c} \quad \text{(cino, c-squark)} \\
b &\rightarrow \hat{b} \quad \text{(bino, b-squark)} \\
t &\rightarrow \hat{t} \quad \text{(tino, t-squark, stop)}
\end{align*}
\]

The spin one-half "inos" are: goldstino \( \hat{G} \) (from the so-called supergoldstone effect), higgsino, \( \hat{H} \), (sometimes called shiggs) and several gauginos:

- photino (another notation \( \hat{\gamma} \))
- gluino (another notation \( \hat{G} \))
- wino (other notations \( \hat{\psi}, \hat{w} \))
- zino (other notations \( \hat{z}, \hat{z} \)).

Of special importance in modern theoretical models is the superpartner of the graviton, the gravitino, particle with spin \( \frac{3}{2} \), often denoted by \( \lambda_{3/2} \).

5.3. Vertices and Coupling Constants

In order to perform simple estimates of cross-sections and decay-rates of processes involving "inos" it is useful to draw corresponding Feynman graphs. Here is a simple recipe, how to draw vertices in the graphs. Consider an ordinary vertex, say, \( \bar{e} \gamma \gamma' \) (Fig. 15a). Draw all lines identically (because in a moment some bosons and fermions will be interchanged). Crown some of the particle pairs with superhats (Fig. 15b). Keep the usual coupling constants. Fig. 15b describes electron-electron-photino vertex with \( e^2 = 4 \pi \alpha \). Some other examples:

1. Quark-quarkino-gluino vertex (Fig. 16), \( g_s^2 = 4 \pi \alpha_s \), where \( g_s \) is the QCD coupling constant.
2. Vertices $\hat{W}e^\nu, \hat{W}e^- e^\nu, \hat{W}e^- e^- (\text{Fig. 17}),$

$$g^2_w = 4\pi \alpha_w = 4\pi \alpha / \sin^2 \theta_w.$$

3. Emission of a goldstino by an electino (Fig. 18).

The coupling constant is dimensionless: $\xi_e = (m_\xi^2 - m_e^2) / m_s^2,$ where $m_s$ is the scale of SUSY breaking. According to the Vogue '83: $m_s \approx 10^{11}$ GeV.

![Fig. 15](image1)

![Fig. 16](image2)

![Fig. 17](image3)

![Fig. 18](image4)

### 2.4. Examples of Reactions and Decays

In order to draw more complex Feynman diagrams, link vertices, keeping inos' lines continuous. Here are some examples:
1. Production of gluinos in gluon-gluon fusion (Fig. 19).
Note that gluino's colour charge is large. Therefore
above the threshold the cross section of the production of
a pair of gluinos will be much larger than that of a pair
of quarks with the same mass:
\[ \sigma(gg \rightarrow \hat{g}\hat{g}) : \sigma(gg \rightarrow q\tilde{q}) = 81:7 \]

\[
\begin{array}{c}
\text{a)} \\
\text{b)}
\end{array}
\]

\[
\begin{array}{c}
\text{c)} \\
\text{d)}
\end{array}
\]

Fig. 19.

2. The decay of gluino into photino and hadrons (Fig. 20):
\[ \hat{g} \rightarrow \tilde{\chi} + q\tilde{\chi} (= \text{hadrons}). \]

\[
\begin{array}{c}
\text{a)} \\
\text{b)}
\end{array}
\]

Fig. 20.

3. Interaction of photino with a quark may result in a produc-
tion of gluino, if the energy of the photino is high enough,
\[ \tilde{\gamma} q \rightarrow \hat{g} q \] (Fig. 21). The corresponding cross section is
\[ \sigma \sim \alpha_s \sin \frac{\Delta m_Q^2}{4}. \]
where \( m_{\tilde{q}} \) is the mass of the (virtual) quarkino.
Another possibility is the elastic scattering (Fig. 22) with $\sigma \sim \alpha^2 \lesssim m_{\tilde{q}}^2$.

The negative results of CHARM beam dump experiment may be interpreted as \textit{\cite{29}}:

$$m_{\tilde{g}} \gtrsim 2 \text{ GeV, } \quad m_{\tilde{q}} \gtrsim 100 \text{ GeV.}$$

It is interesting to search for gluinos at SPS p\bar{p}-collider. If gluinos are heavy their signature is large missing $p_T$ without accompanying leptons.

4. Decays of $W$ (and $Z$) intoinos, if the inos are light (Fig. 23).

5. Decays of $\hat{W}$ and $\hat{Z}$ (Fig. 24)

$$\hat{W} \rightarrow \hat{\gamma} + \text{ hadrons, } \quad \hat{W} \rightarrow \hat{\chi} + \text{ hadrons, } \quad \hat{Z} \rightarrow \hat{\gamma} + \text{ hadrons.}$$
6. Decays of $\tilde{\nu}$:
- $\tilde{\nu} \rightarrow \tilde{\nu} v$ (Fig. 25),
- $\tilde{\nu} \rightarrow \nu \tilde{\nu}$ (Fig. 26),
- $\tilde{\nu} \rightarrow \tilde{\nu} e^- + \text{hadrons}$ (Fig. 27). Let us mention that decay rate of the first of these processes (muino $\rightarrow$ goldstino + neutrino) is of the order of $\frac{m_{\tilde{\nu}}^3}{m_\nu^4}$.

5.5. Examples of Mass Formulas

Nobody really knows what the masses of ino's are. Here are predictions from some of the recent preprints:
All these formulas contain $m_{3/2}$, the mass of gravitino. An educated guess:

$$300 \text{ GeV} \lesssim m_{3/2} \lesssim 50 \text{ GeV}.$$ 

Note that the mixing between higgsinos ($H$) and weakinos ($\tilde{W}$ and $\tilde{Z}$) is essential:

- The mixing $H^+ \leftrightarrow \tilde{W}^+$ produces higgwinos,
- The mixing $H^0 \leftrightarrow \tilde{Z}^0$ produces higgzinos.

Some of the ino's could be very light:

$$m_{\tilde{g}} \gtrsim 2 \text{ GeV}, \quad m_{\tilde{\chi}} \gtrsim 0.2 \text{ GeV}.$$

6. On the Role of Gravitino in the Local $N=1$ SUSY (Supergravity)

The interaction of transversal components of gravitinos is extremely weak, like that of graviton, the vertices being proportional to $E/mL$ where $E$ is the energy of gravitino.

The vertices for the emission and absorption of gravitino's longitudinal component is proportional to $E^2/mL$ and may be observable if $E/mL \gg 1$. In this case they appear in the form of goldstinos, $G$.

An assumption that $m_{3/2} \sim mL$ contradicts the mass formulas discussed above, because in this case inos are almost degenerate with their ordinary counterparts: $m_\chi \approx m_e$, etc.

Another assumption, $m_{3/2} \sim mL$, seems OK at present. With $m_\chi^2 = m_{3/2} mL$ we have $m_\chi \approx 10^{11} \text{ GeV}$. In this case both transversal and longitudinal components of gravitino are practically unobservable at our energies. And in spite of this, gravitino plays a very important role!

This lecture is based on several recent preprints $^{30-33}$ which are but a drop from an ocean. For further information see lectures by prof. G. Ross.

Die Welt ist ein vierdimensionales Kontinuum und läßt sich deshalb auf vier Koordinaten $x_0, x_1, x_2, x_3$ beziehen. Der Übergang zu einem anderen Koordinatensystem $\xi_i$ wird durch stetige Transformationsformeln

\begin{equation}
\xi_i = \hat{f}_i (\bar{x}_0, \bar{x}_1, \bar{x}_2, \bar{x}_3) \quad (i = 0, 1, 2, 3)
\end{equation}

vermittelt. An sich ist unter den verschiedenen möglichen Koordinatensystemen keines ausgezeichnet. Die Relativkoordinaten $dx_i$ eines zu dem Punkte $P = (x_i)$ unendlich benachbarten $P' = (x_i + dx_i)$ sind die Komponenten der infinitesimalen Verschiebung $\overrightarrow{PP'}$ eines ,,Linienelements" in $P$. Sie transformieren sich beim Übergang (1) zu einem anderen Koordinatensystem $\bar{x}_i$ linear:

\begin{equation}
dx_i = \sum_k \alpha_k^i \, d\bar{x}_k;
\end{equation}

$\alpha_k^i$ sind die Werte der Ableitungen $\partial \hat{f}_i / \partial \bar{x}_k$ im Punkte $P$. In der gleichen Weise transformieren sich die Komponenten $\xi^i$ irgendeines Vektors in $P$. Mit einem die Umgebung von $P$ bedeckenden Koordinatensystem ist ein ,,Achsenkreuz" in $P$ verknüpft, bestehend aus den ,,Einheitsvektoren" $\epsilon_i$ mit den Komponenten $\delta^0_i, \delta^1_i, \delta^2_i, \delta^3_i$:

\[ \delta^{i_k}_i = \begin{cases} 0 & (i \neq k) \\ 1 & (i = k) \end{cases} \]
An seine Stelle aber trat bei Berücksichtigung der Gravitation der Gegensatz von elektromagnetischem Feld („Materie im weiteren Sinne“, wie Einstein sagt) und Gravitationsfeld; er zeigt sich am deutlichsten in der Zweiteilung der Hamiltonschen Funktion, welche der Einsteinschen Theorie zugrunde liegt.\(^1\) Auch dieser Zwiespalt wird durch unsere Theorie überwunden. Der Integrand der Wirkungsgröße \(\int W \, dx\) muß eine aus der Metrik entspringende skalare Dichte \(W\) sein, und die Naturgesetze sind zusammengefaßt in dem Hamiltonschen Prinzip: Für jede infinitesimale Änderung \(\delta\) der Weltmetrik, die außerhalb eines endlichen Bereichs verschwindet, ist die Änderung

\[
\delta \int W \, dx = \int \delta W \, dx
\]

der gesamten Wirkungsgröße \(= 0\) (die Integrale erstrecken sich über die ganze Welt oder, was auf dasselbe hinauskommt, über einen endlichen Bereich, außerhalb dessen die Variation \(\delta\) verschwindet). Die Wirkungsgröße ist in unserer Theorie notwendig eine reine Zahl; anders kann es ja auch nicht sein, wenn ein Wirkungsquantum existieren soll. Von \(W\) werden wir annehmen, daß es ein Ausdruck 2. Ordnung ist, d. h. aufgebaut ist einerseits aus den \(g_{ik}\) und deren Ableitungen 1. und 2. Ordnung, andererseits aus den \(\varphi_i\) und deren Ableitungen 1. Ordnung.

Das einfachste Beispiel ist die Maxwellscbe Wirkungsdichte \(L\). Wir wollen aber in diesem Kapitel keinen speziellen Ansatz für \(W\) zugrunde legen, sondern untersuchen, was sich allein aus dem Umstande erschließen läßt, daß \(\int W \, dx\) ein koordinaten- und eichinvariantes Integral ist. Wir bedienen uns dabei einer von F. Klein angegebenen Methode.\(^2\)

Folgerungen aus der Invarianz der Wirkungsgröße. a) Eichinvarianz. Erteilen wir dem die Metrik relativ zu einem Bezugssystem beschreibenden Größen \(\varphi_i\), \(g_{ik}\) beliebige unendlich kleine Zuwächse \(\delta \varphi_i\), \(\delta g_{ik}\) und bedeutet \(\mathcal{X}\) ein endliches Weltgebiet, so ist es der Effekt der partiellen Integration, daß das Integral der zugehörigen Änderung \(\delta W\) von \(W\) über das Gebiet \(\mathcal{X}\) in zwei Teile zerlegt wird: ein Divergenzintegral und ein

---

Eine neue Erweiterung der Relativitätstheorie.

Integral, dessen Integrand nur noch eine lineare Kombination von $\delta \varphi_i$ und $\delta g_{ik}$ ist:

$$\int \frac{\partial (\delta \mathbf{v}_k)}{\partial x_i} \, dx = \int \left( \delta \varphi_i + \frac{1}{2} \mathbf{M}^{ik} \delta g_{ik} \right) \, dx.$$  \hspace{1cm} (8)

Dabei sind $\mathbf{v}_i$, $\delta \mathbf{v}_i$ die Komponenten je einer kontravarianten Vektordichte, $\mathbf{M}^i_k$ aber die einer gemischten Tensordichte 2. Stufe (im eigentlichen Sinne). Die Komponenten $\delta \mathbf{v}_i$ sind lineare Kombinationen von $\delta \varphi_i$, $\delta g_{ik}$ und $\delta g_{ik, r}$, $\left\{ \delta g_{ik, r} = \frac{\partial g_{ik}}{\partial x_r} \right\}$.

Wir drücken jetzt zunächst aus, daß $\int \delta \mathbf{v}_i \, dx$ sich nicht ändert, wenn die Eichung der Welt infinitesimal abgeändert wird. Ist $a = 1 + \pi$ das Eichverhältnis zwischen ursprünglicher und abgeänderter Eichung, so ist $\pi$ ein den Vorgang charakterisierendes infinitesimal skalarfeld, das willkürlich vorgegeben werden kann. Bei diesem Prozefßerfahren die Fundamentalgrößen die Zuwächse

$$\delta g_{ik} = \pi \cdot g_{ik}, \quad \delta \varphi_i = - \frac{\partial n}{\partial x_i}.$$  \hspace{1cm} (9)

Substituiert wir diese Werte in $\delta \mathbf{v}_i$, so mögen die Ausdrücke

$$\delta \mathbf{v}_i = \frac{\partial n}{\partial x_i} \cdot \mathbf{v}_i$$  \hspace{1cm} (10)

hervorgehen. Die Variation (8) des Wirkungsintegrals muß für (9) verschwinden: so formulieren wir die Tatsache der Eichinvarianz.

$$\int \frac{\partial \delta \mathbf{v}_i}{\partial x_i} \, dx + \int \left( - \mathbf{v}_i \frac{\partial n}{\partial x_i} + \frac{1}{2} \mathbf{M}^i_k \cdot \pi \right) \, dx = 0.$$  \hspace{1cm} (11)

Formt man den ersten Term des zweiten Integrals noch durch partielle Integration um, so kann man statt dessen schreiben:

$$\int \frac{\partial (\delta \mathbf{v}_i - \pi \mathbf{w}_i)}{\partial x_i} \, dx + \int \pi \left( \frac{\partial \mathbf{w}_i}{\partial x_i} + \frac{1}{2} \mathbf{M}^i_k \right) \, dx = 0.$$  \hspace{1cm} (12)

Daraus ergibt sich zunächst die Identität

$$\frac{\partial \mathbf{w}_i}{\partial x_i} + \frac{1}{2} \mathbf{M}^i_k = 0$$  \hspace{1cm} (12)

in der aus der Variationsrechnung bekannten Weise: wäre diese Ortsfunktion an einer Stelle ($x_i$) von 0 verschieden, etwa
Quantentheorie und fünfdimensionale Relativitätstheorie.

Von Oskar Klein in Kopenhagen.

(Eingegangen am 28. April 1926.)

Auf den folgenden Seiten möchte ich auf einen einfachen Zusammenhang hinweisen zwischen der von Kaluzza\(^1\) vorgeschlagenen Theorie für den Zusammenhang zwischen Elektrizitäts- und Gravitationselektrizitätsregel und der von de Bröl\(^{2}\) und Schrödinger\(^3\) angegebenen Methode zur Behandlung der Quantenprobleme andererseits. Die Theorie von Kaluzza geht darauf hinaus, dass die zehn Einstein'schen Gravitationspotentiale \(g_{ik}\) und die vier elektromagnetischen Potentiale \(q_i\) in Zusammenhang zu bringen mit den Koeffizienten \(y_{ik}\) eines Linienelements von einem vierdimensionalen Raum, der außer den vier gewöhnlichen Dimensionen noch eine fünfte Dimension enthält. Die Bewegungsgleichungen der elektromagnetischen Teilchen nehmen hierbei auch in elektromagnetischen Feldern die Gestalt von Gleichungen geodätischer Linien an. Wenn dieselben als Stromgleichungen gedacht werden, indem die Materie als eine Art Wellenausbreitung betrachtet wird, kommt man fast von selbst zu einer partiellen Differentialgleichung zweiter Ordnung, die als eine Verallgemeinerung der gewöhnlichen Wellengleichung angesehen werden kann. Werden nun solche Lösungen dieser Gleichung betrachtet, bei denen die fünfte Dimension rein harmonisch umtritt mit einer bestimmten mit der Planckschen Konstante zusammenhängenden Periode, so kommt man eben zu den oben erwähnten quantentheoretischen Methoden.

§ 1. Fünfdimensionale Relativitätstheorie. Ich fange damit an, eine kurze Darstellung von der fünfdimensionalen Relativitätstheorie zu geben, die sich nahe an die Theorie von Kaluzza anschließt, aber in einigen Punkten von derselben abweicht.

Betrachten wir ein fünfdimensionales Riemannsches Linienelement, für welches wir einen vom Koordinatensystem unabhängigen Sinus postulieren. Wir schreiben dasselbe:

\[
d\sigma = \sqrt{\sum' y_{ik} dx^i dx^k}, \tag{1}\]

wo das Zeichen \(\sum'\), wie überall im folgenden, eine Summation über die doppelt vorkommenden Indizes von 0 bis 4 angibt. Hierbei bezeichnen \(x^0 \ldots x^4\) die fünf Koordinaten des Raumes. Die 15 Größen \(y_{ik}\) sind die kovarianten Komponenten eines fünfdimensionalen symmetrischen Tensors. Um von derselben zu den Größen \(g_{ik}\) und \(q_i\) der gewöhnlichen Relativitätstheorie zu kommen, müssen wir gewisse spezielle Annahmen machen. Erstens müssen vier der Koordinaten, sagen wir \(x^1, x^2, x^3, x^4\), stets den gewöhnlichen Zeiteinheiten charakterisieren. Zweitens dürfen die Größen

---


\(3\) E. Schrödinger, Ann. d. Phys. 79, 361 und 389, 1925.

Zschrift fur Physik. Bd. XXXVII.
sich die Wellen nach den Gesetzen der geometrischen Optik ausbreiten. Es mag noch hervorgehoben werden, daß wegen (42) die Beziehungen (44), (45) bei den Koordinatentransformationen (2) invariant bleiben.

Betrachten wir nun auch die Gleichung (24) in dem Falle, wo $\omega$ nicht so groß ist, daß wir nur die in $\omega$ quadratischen Glieder zu berücksichtigen brauchen. Wir beschränken uns dabei auf den einfachen Fall eines elektrostatischen Feldes. Dann haben wir in kartesischen Koordinaten:

\begin{align*}
\dot{\theta} &= \dot{x} - e V t, \\
\dot{s} &= \dot{x}^2 + \dot{y}^2 + \dot{z}^2 - e^2 dt^2.
\end{align*}

(46)

Also ergibt sich:

$$H = \frac{1}{2} (p_x^2 + p_y^2 + p_z^2) - \frac{1}{2} \left( \mu_t + e V \mu_0 \right)^2 = \frac{m^2 c^2}{2} p_n. \quad (47)$$

In der Gleichung (24) können wir nun die mit $\frac{\mu_1}{r}$ proportionalen Größen vernachlässigen, denn die Dreindizessymbolden sind in diesem Falle nach (17) keine mit der Gravitationskonstante $\kappa$ proportionale Größen. Wir bekommen also:

$$\Delta \psi + \frac{4 \pi^2}{r^3 h^2} \left[ (\hbar V - e V) - m^2 c^4 \right] \psi = 0. \quad (50)$$

Setzen wir noch:

$$\hbar V = m c^3 \mp E,$$

so bekommen wir die von Schrödinger gegebene Gleichung, deren stehende Schwingungen bekanntlich mit $E$ entsprechen, die mit


2) Schrödinger, l. c.
den aus der Heisenbergschen Quantentheorie berechneten Energie-

werten identisch sind. Wie man sieht, ist \( E \) in dem Grenzfall der ge-

ometrischen Optik gleich der auf gewöhnliche Weise definierten mecha-
nischen Energie. Die Frequenzbedingung besagt, wie Schrödinger 

hervorgehoben hat nach (51), daß die zu dem System gehörenden Licht-

frequenzen den aus den unterschiedlichen Werten der Frequenz \( \nu \) gebildeten 

differenzen gleich sind.

§ 3. Schlußbemerkungen. Wie die Arbeiten von de Broglie 

sind obenstehende Überlegungen aus dem Bestreben entstanden, die 

Analogie zwischen Mechanik und Optik, die in der Hamiltonschen 

Methode zum Vorschein kommt, für ein tieferes Verständnis der quanten-

erscheinungen auszunutzen. Daß dieser Analogie ein reeller physi-

kalischer Sinn zukommt, scheint ja die Ähnlichkeit der Bedingungen für 

die stationären Zustände von Atomystemen mit den Interferenz-

erscheinungen der Optik anzudeuten. Nun stehen bekanntlich Begriffe 

wie Punktladung und materieller Punkt schon der klassischen Feld-

physik fremd gegenüber. Auch wurde ja öfters die Hypothese aus-

gesprochen, daß die materiellen Teilchen als spezielle Lösungen der 

Feldgleichungen aufzufassen sind, welche das Gravitationsfeld und das-

elektromagnetische Feld bestimmen. Es liegt nahe, die genannte Ana-

alogie zu dieser Vorstellung in Beziehung zu bringen. Denn nach dieser 

Hypothese ist es ja nicht so befremdend, daß die Bewegung der mate-

riellen Teilchen Ähnlichkeiten aufweist mit der Ausbreitung von Wellen.

Die in Rede stehende Analogie ist jedoch unvollständig, solange man 
eine Wellenausbreitung in einem Raum von nur vier Dimensionen be-

trachtet. Dies kommt schon in der variablen Geschwindigkeit der 

materiellen Teilchen zum Vorschein. Denkt man sich aber die beob-
auchte Bewegung als eine Art Projektion auf den Zeitraum von einer 

Wellenausbreitung, die in einem Raum von fünf Dimensionen stattfindet, 

so läßt sich, wie wir sahen, die Analogie vollständig machen. Mathematisch ausgedrückt heißt dies, daß die Hamilton-Jacobi-gleichung 

nicht als Charakteristikengleichung einer vierdimensionalen, wohl aber einer fünfdimensionalen Wellengleichung aufgefaßt werden kann. 

In dieser Weise wird man zu der Theorie von Kaluza geführt.

Obwohl die Einführung einer fünften Dimension in unsere physi-

kalischen Betrachtungen von vornherein befremdend sein mag, wird eine 

radikale Modifikation der den Feldgleichungen zugrunde gelegten 

Geometrie doch wieder in ganz anderer Weise durch die Quantentheorie 

nahegelegt. Denn es ist bekanntlich immer weniger wahrscheinlich
Oskar Klein, Quantentheorie und fünfdimensionale Relativitätstheorie.

geworden, daß die Quantenerscheinungen eine einheitliche raumzeitliche Beschreibung zulassen, wogegen die Möglichkeit, diese Erscheinungen durch ein System von fünfdimensionalen Feldgleichungen darzustellen, wohl nicht von vornherein auszuschließen ist 1). Ob hinter diesen An- deutungen von Möglichkeiten etwas Wirkliches besteht, muß natürlich die Zukunft entscheiden. Jedenfalls muß betont werden, daß die in dieser Note versuchte Behandlungsweise, sowohl was die Feldgleichungen als auch die Theorie der stationären Zustände betrifft, als ganz provisorisch zu betrachten ist. Dies kommt wohl besonders in der auf S. 898 erwähnten schematischen Behandlungsweise der Materie zum Vorschein, sowie in dem Umstand, daß die zwei Arten von elektrischen Teilchen durch verschiedene Gleichungen vom Schrödingerschen Typus behandelt werden. Auch wird die Frage ganz offen gelassen, ob man sich bei der Beschreibung der physikalischen Vorgänge mit den 14 Potentialen be- gnügt oder ob die Schrödingersche Methode die Einführung einer neuen Zustandsgröße bedeutet.

Mit den in dieser Note mitgeteilten Überlegungen habe ich mich sowohl in dem Physikalischen Institut der University of Michigan, Ann Arbor, wie in dem hiesigen Institut für theoretische Physik beschäftigt. Ich möchte auch an dieser Stelle Prof. H. M. Randall und Prof. N. Bohr meinen wärmsten Dank aussprechen.

1) Bemerkungen dieser Art, die Prof. Bohr bei mehreren Gelegenheiten gemacht hat, haben einen entschiedenen Einfluß auf das Entstehen der vorliegenden Note gehabt.
Über
die invariante Form der Wellen- und der Bewegungsgleichungen für einen geladenen Massenpunkt.

Von V. Fock in Leningrad.

(Eingegangen am 30. Juli 1926.)

Die Schrödingersche Wellengleichung wird als invariante Laplacesche Gleichung und die Bewegungsgleichungen als diejenigen einer geodätischen Linie im fünfdimensionalen Raum geschrieben. Der überzählige fünfte Koordinatenparameter steht in enger Beziehung zu der linearen Differentialform der elektromagnetischen Potentiale.


1. Spezielle Relativitätstheorie.

Die Lagrangesche Funktion für die Bewegung eines geladenen Massenpunktes ist, in leicht verständlicher Bezeichnungsweise,

\[ L = -mc^2 \left( 1 - \frac{v^2}{c^2} \right) \frac{1}{\sqrt{\mathfrak{A} \cdot v - e \varphi}} \]  

(1)

und die entsprechende Hamilton-Jacobische Gleichung (II. P.) lautet:

\[ (\nabla W)^2 - \frac{1}{c^2} \left( \frac{\partial W}{\partial t} \right)^2 - \frac{2e}{c} \left( \mathfrak{A} \cdot \nabla W + \frac{e}{c} \frac{\partial W}{\partial t} \right) \]

\[ = m^2 c^2 + \frac{e^2}{c^2} \left( \mathfrak{A}^2 - q^2 \right) = 0. \]  

(2)

1) Die Idee dieser Arbeit ist in einem Gespräch mit Prof. V. Fock entstanden, dem ich auch manche wertvolle Ratschläge verdanke.


2) Der Verfasser hat mir liebenswürdigerweise die Möglichkeit gegeben, seine Arbeit im Manuskript zu lesen.
V. Fock, Über die invariante Form der Wellengleichungen usw.

Analog dem in unserer früheren Arbeit\(^1\) gebrauchten Ansätze setzen wir hier

\[
\text{grad } W = \frac{\text{grad } \psi}{\partial \psi \partial p}; \quad \frac{\partial W}{\partial t} = \frac{\partial \psi}{\partial \psi \partial p},
\]

(3)

wo \( p \) einen neuen Parameter von der Dimension des Wirkungsquantums bezeichnet. Nach Multiplikation mit \( \left( \frac{\partial \psi}{\partial p} \right)^2 \) erhalten wir eine quadratische Form

\[
Q = (\text{grad } \psi)^2 - \frac{1}{c^4} \left( \frac{\partial \psi}{\partial t} \right)^2 - \frac{2e}{c} \frac{\partial \psi}{\partial p} \left( \mathfrak{u}_0 \cdot \text{grad } \psi + \frac{q}{c} \frac{\partial \psi}{\partial t} \right) + \left[ \mathfrak{u}^2 c^2 + \frac{c^2}{e^2} (\mathfrak{u}^2 - q^2) \right] \left( \frac{\partial \psi}{\partial p} \right)^2.
\]

(4)

Wir bemerken, daß die Koeffizienten der nullten, ersten und zweiten Potenz von \( \frac{\partial \psi}{\partial p} \) vierdimensionale Invarianten sind. Ferner bleibt die Form \( Q \) invariant, wenn man

\[
\begin{align*}
\mathfrak{u} &= \mathfrak{u}_1 + \text{grad } \xi, \\
\varphi &= \varphi_1 - \frac{1}{c} \frac{\partial \xi}{\partial t}, \\
\nu &= \nu_1 - \frac{e}{c} \xi
\end{align*}
\]

(5)

setzt, wo \( \xi \) eine willkürliche Funktion der Koordinaten und der Zeit bezeichnet. Die letztere Transformation läßt auch die lineare Differentialform

\[
d' \Omega = \frac{e}{mc^2} (\mathfrak{u}_x dx + \mathfrak{u}_y dy + \mathfrak{u}_z dz) - \frac{e}{mc} \varphi dt + \frac{1}{mc} dp
\]

(6)

invariant\(^2\).

Wir wollen nun die Form \( Q \) als das Quadrat des Gradienten der Funktion \( \psi \) im fünfdimensionalen Raum \( (R^5) \) auffassen, und suchen das entsprechende Linienelement. Man findet leicht

\[
d's^2 = ds^2 + df^2 - c^2 dt^2 + (d' \Omega)^2.
\]

(7)

\(^1\) V. Fock, Zur Schrödingerischen Wellenmechanik, ZS. f. Phys. 38, 6.

\(^2\) Das Zeichen \( d' \) soll andeuten, daß \( d' \Omega \) kein vollständiges Differential ist.
Die Laplacesche Gleichung in \( R_4 \) lautet:

\[
\Delta \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 2 \epsilon \left( \varepsilon \text{ grad} \frac{\partial \psi}{\partial t} + \frac{\psi}{c} \frac{\partial^2 \psi}{\partial t \partial p} \right)
\]

\[
- \frac{\epsilon}{c} \frac{\partial \psi}{\partial p} \left( \text{div} \varepsilon + \frac{1}{c} \frac{\partial \psi}{\partial t} \right) + \left[ m^2 c^2 + \frac{\epsilon^2}{c^2} (\varepsilon^2 - \psi^2) \right] \frac{\partial^2 \psi}{\partial p^2} = 0. \tag{8}
\]

Sie bleibt, ebenso wie (7) und (4), bei Lorentztransformationen und bei Transformationen (5) invariant.

Da die Koeffizienten der Gleichung (8) den Parameter \( p \) nicht enthalten, können wir die Abhängigkeit der Funktion \( \psi \) von \( p \) in Form eines Exponentialfaktors ansetzen, und zwar müssen wir, um Übereinstimmung mit der Erfahrung zu erhalten

\[
\psi = \psi_0 e^{\frac{2 \pi i p}{\hbar}} \tag{9}
\]

setzen \(^1\). Die Gleichung für \( \psi_0 \) ist gegenüber Lorentztransformationen, nicht aber gegenüber Transformationen (5), invariant. Die Bedeutung des überzähligen Koordinatenparameters \( p \) scheint nämlich gerade darin zu liegen, daß er die Invarianz der Gleichungen in bezug auf die Addition eines beliebigen Gradienten zum Viererpotential bewirkt.

Es sei hier bemerkt, daß die Koeffizienten der Gleichung für \( \psi_0 \) im allgemeinen komplex sind.

Nimmt man ferner an, daß diese Koeffizienten von \( t \) nicht abhängen, und setzt man

\[
\psi_0 = e^{-\frac{2 \pi i}{\hbar} (E + m c^2) t} \psi_0 \tag{10}
\]

so erhält man für \( \psi_1 \) eine zeitfreie Gleichung, welche mit der in unserer früheren Arbeit aufgestellten Verallgemeinerung der Schrödingerschen Wellengleichung identisch ist. Diejenigen Werte von \( E \), für welche eine Funktion \( \psi_1 \) existiert, die gewissen Endlichkeits- und Stetigkeitsforderungen genügt, sind dann die Bohrschen Energienebenen. Aus den soeben angeführten Betrachtungen folgt, daß die Addition eines Gradienten zum Viererpotential keinen Einfluß auf die Energienebenen ausüben kann. Die beiden mit den Vektorpotentialen \( \varepsilon \) und \( \varepsilon = \varepsilon - \text{grad} f \) erhaltenen Funktionen \( \psi_1 \) und \( \overline{\psi}_1 \) würden sich nämlich nur um einen Faktor \( e^{\frac{2 \pi i}{\hbar} f} \) vom absoluten Betrage \( 1 \) unterscheiden und folglich (bei sehr allgemeinen Voraussetzungen über die Funktion \( f \)) die gleichen Stetigkeits-eigenschaften haben.

\(^1\) Das Auftreten des mit der Linearform verbundenen Parameters \( p \) in einer Exponentialfunktion könnte vielleicht mit einigen von E. Schrödinger (ZS. f. Phys. 12, 13 [1923] bemerkten Beziehungen im Zusammenhang stehen.
2. Allgemeine Relativitätstheorie.

A. Wellengleichung. Für das Linienelement im fünfdimensionalen Raum setzen wir an:

\[ ds^2 = \sum_{i,k=1}^{5} \gamma_{ik} dx_i dx_k = \sum_{i,k=1}^{5} g_{ik} dx_i dx_k + \frac{c^2}{m^2} \left( \sum_{i=1}^{5} u_i \right)^2 \]  

Hier sind die Größen \( g_{ik} \) die Komponenten des Einsteinschen Fundamentaltensors, die Größen \( u_i \) \((i = 1, 2, 3, 4)\) die durch \( c^2 \) dividierten Komponenten des Viererpotentials, also

\[ \sum_{i=1}^{5} u_i dx_i = \frac{1}{c^2} (\mathcal{A}_x dx + \mathcal{A}_y dy + \mathcal{A}_z dz - u_i \, dt), \]

die Größe \( u_i \) eine Konstante und \( x_5 \) der überzählige Koordinatenparameter. Alle Koeffizienten sind reell und von \( x_5 \) unabhängig.

Die Größen \( g_{ik} \) und \( u_i \) hängen nur vom Felde, nicht aber von der Beschaffenheit des Massenpunktes ab; die letztere wird durch den Faktor \( \frac{c^2}{m^2} \) repräsentiert. Zur Abkürzung wollen wir aber die von \( \frac{c}{m} \) abhängigen Größen

\[ \frac{c}{m} \; u_i = a_i \quad (i = 1, 2, 3, 4, 5) \]
einführen und folgende Verabredung treffen: bei der Summation von 1 bis 5 wird das Summenzeichen angeschrieben, bei der Summation von 1 bis 4 dagegen unterdrückt.

Mit diesen Bezeichnungen finden wir

\[ \gamma_{ik} = g_{ik} + u_i a_k; \quad g_{ik} = 0 \]

\[ \gamma = \| \gamma_{ik} \| = a_i^2 g \]  

\[ \gamma^{ik} = g^{ik} \]

\[ \gamma^{ik} = - \frac{1}{u_i} \; g^{ik} a_i = - \frac{a_i}{u_i} \]  

\[ \gamma^{55} = \frac{1}{(a_i)^2} \cdot (1 + u_i a_i) \]

Die der Gleichung (8) entsprechende Wellengleichung lautet:

\[ \sum_{i,k=1}^{5} \frac{\partial}{\partial x_i} \left( \sqrt{-\gamma} \; \gamma^{ik} \frac{\partial u_i}{\partial x_k} \right) = 0 \]  

48
oder, ausführlicher geschrieben,

\[
\frac{1}{\sqrt{-g}} \frac{\partial}{\partial x_i} \left( \sqrt{-g} g^{ik} \frac{\partial \psi}{\partial x_k} \right) - \frac{2}{a_5} a^i \frac{\partial^2 \psi}{\partial x_i \partial x_b} + \frac{1}{(a_6)^2} (1 + a_i a^i) \frac{\partial^2 \psi}{\partial x_b^2} = 0. \tag{18}
\]

Führt man endlich die Funktion \( \psi_0 \) und die Potentiale \( q_i \) ein, so läßt sich diese Gleichung schreiben:

\[
\frac{1}{\sqrt{-g}} \frac{\partial}{\partial x_i} \left( \sqrt{-g} g^{ik} \frac{\partial \psi_0}{\partial x_k} \right) - \frac{4 \pi}{\hbar} \sqrt{-1} e a^i \frac{\partial \psi_0}{\partial x_i} - \frac{4 \pi^2 e^2}{\hbar^2} (m^2 + e^2 q_i q') \psi_0 = 0. \tag{19}
\]

B. Bewegungsgleichungen. Wir wollen nun die Bewegungsgleichungen eines geladenen Massenpunktes als diejenigen einer geodätischen Linie in \( \mathbb{R}_5 \) aufstellen.

Zu diesem Zwecke müssen wir zunächst die Christoffelschen Klammersymbole berechnen. Wir bezeichnen die fünfdimensionalen Klammersymbole mit \( \{ k l \}_{r 5} \) und die vierdimensionalen mit \( \{ k l \}_{r 4} \). Wir führen ferner die kovariante Ableitung des Viererpotentials ein:

\[
A_{lk} = \frac{\partial a_l}{\partial x_k} - \{ k l \}_{r 4} a_r, \tag{20}
\]

und spalten den Tensor \( 2 A_{lk} \) in seinen symmetrischen und antisymmetrischen Teil:

\[
B_{lk} = A_{lk} + A_{kl}, \quad M_{lk} = A_{lk} - A_{kl} = \frac{\partial a_l}{\partial x_k} - \frac{\partial a_k}{\partial x_l}. \tag{21}
\]

Wir haben dann

\[
\begin{align*}
\{ k l \}_{r 5} &= \{ k l \}_{r 4} + \frac{1}{2} (a_k g^{ir} M_{il} + a_l g^{ir} M_{ik}), \\
\{ k l \}_{5 b} &= \frac{1}{2 a_5} B_{il} - \frac{1}{2 a_5} (a_k a^i M_{il} + a_l a^i M_{ik}), \\
\{ k 5 \}_{5 b} &= - \frac{1}{2} a^i M_{ik}, \\
\{ 5 5 \}_{5 b} &= 0, \\
\{ 5 5 \}_{b} &= 0.
\end{align*} \tag{22}
\]
Über die invariante Form der Wellen- und der Bewegungsgleichungen usw.

Die Gleichungen der geodätischen Linie in $\mathbb{R}_6$ lauten dann:

$$\frac{d^2 x_r}{ds^2} + \left\{ \begin{array}{l} k_l \\ r \end{array} \right\} \frac{dx_k}{ds} \frac{dx_l}{ds} + \frac{d\Omega}{ds} \cdot g^{ir} M_{il} \frac{dx_l}{ds} = 0, \quad (23)$$

$$\frac{d^2 x_b}{ds^2} + \frac{1}{2} B_{ik} \frac{dx_k}{ds} \frac{dx_l}{ds} - \frac{1}{a_o} \frac{d\Omega}{ds} a^l M_{il} \frac{dx_l}{ds} = 0. \quad (24)$$

Hier bezeichnet $d\Omega$, wie früher, die Linearform

$$d\Omega = a_i dx_i + a_o dx, \quad (25)$$

Multipliziert man die vier Gleichungen (23) mit $a_i$, die fünfte (24) mit $a_o$ und addiert, so erhält man eine Gleichung, die man in der Form

$$\frac{d}{ds} \left( \frac{d\Omega}{ds} \right) = 0 \quad (26)$$

schreiben kann. Es gilt also

$$\frac{d\Omega}{ds} = \text{const} \quad (27)$$

Multipliziert man (23) mit $g_{ra} \frac{dx_a}{ds}$ und summiert über $r$ und $a$, so folgt wegen der Antisymmetrie von $M_{ik}$

$$\frac{d}{ds} \left( g_{ra} \frac{dx_r}{ds} \frac{dx_a}{ds} \right) = 0 \quad (28)$$

oder, wenn man die Eigenzeit $\tau$ durch die Formel

$$g_{ik} \frac{dx_l}{ds} \frac{dx_k}{ds} = -c^2 d\tau^2 \quad (29)$$

einführt,

$$\frac{d}{ds} \left( \frac{d\tau}{ds} \right)^2 = 0. \quad (30)$$

Die Gleichung (28) oder (30) ist übrigens wegen

$$ds^2 = -c^2 d\tau^2 + (d\Omega)^2 \quad (31)$$

eine Folge von (26).

Aus dem Gesagten folgt, daß die Gleichung (24) eine Folge von (23) ist; wir können sie also beiseite lassen. Führen wir in (23) die Eigenzeit als unabhängige Veränderliche ein, so kommt der fünfte Parameter ganz zum Fortfall; wir unterdrücken noch das Zeichen $4$ am Klammersymbol:

$$\frac{d^3 x_r}{d\tau^3} + \left\{ \begin{array}{l} k_l \\ r \end{array} \right\} \frac{dx_k}{d\tau} \frac{dx_l}{d\tau} + \frac{d\Omega}{d\tau} g^{ir} M_{il} \frac{dx_l}{d\tau} = 0. \quad (32)$$

Das letzte Glied auf der linken Seite stellt die Lorentzsche Kraft dar. In der speziellen Relativitätstheorie läßt sich die erste dieser Gleichungen schreiben

$$m \frac{d}{dt} \frac{dx}{d\tau} + \frac{1}{c} \frac{d\Omega}{d\tau} \left[ e \left( \dot{z} H_y - \dot{y} H_z + \frac{\partial \varphi}{\partial t} \right) + e \frac{\partial \varphi}{\partial x} \right] = 0. \quad (33)$$
Um Übereinstimmung mit der Erfahrung zu erhalten, muß der Faktor der eckigen Klammer den Wert 1 haben. Es gilt also
\[ \frac{d' \mathcal{Q}}{d \tau} = c \]
und
\[ ds^2 = 0. \]
Die Bahnen des Massenpunktes sind also geodätische Nulllinien im fünfdimensionalen Raum.

Um die Hamilton-Jacobische Gleichung zu erhalten, setzen wir das Quadrat des fünfdimensionalen Gradienten einer Funktion \( \psi \) gleich Null.
\[ g^{tk} \frac{\partial \psi}{\partial x_t} \frac{\partial \psi}{\partial x_k} - 2 \frac{\partial \psi}{\partial x_0} \frac{a_i}{a_0} \frac{\partial \psi}{\partial x_i} + (1 + a_i a^i) \left( \frac{1}{a_0} \frac{\partial \psi}{\partial x_0} \right)^2 = 0. \quad (36) \]
Setzen wir hier
\[ m c a_0 \frac{\partial \psi}{\partial x_i} = \frac{\partial W}{\partial x_i}, \quad (37) \]
und führen wir statt \( a_i \) die Potentiale \( q_i \) ein, so erhalten wir eine Gleichung
\[ g^{tk} \frac{\partial W}{\partial x_t} \frac{\partial W}{\partial x_k} - 2 e c q^t \frac{\partial W}{\partial x_t} + e^2 (m^2 + e^2 q_t q^t) = 0, \quad (38) \]
die als Verallgemeinerung unserer Gleichung (2), die uns als Ausgangspunkt diente, gelten kann.

Leningrad, Physikalisches Institut der Universität, 24. Juli 1926.
Quantenmechanische Deutung der Theorie von Weyl 1).


(Eingegangen am 25. Februar 1927.)

Kap. II. Die Undulationsmechanik von de Broglie und die Theorie von Weyl.
   § 1. Die Identität von ψ und Weyls Eichstrecke.
   § 2. Nichtintegrabilität schließt Eindeutigkeit nicht aus.
Kap. III. Quantenmechanische Umdeutung der Theorie von Weyl.

Kapitel I. Die Theorie von Weyl.

Die Idee einer „reinen Nahgeometrie“, zuerst von Riemann konzipiert, hat bekanntlich kürzlich durch Weyl eine außerordentlich schöne und einfache Vervollständigung erfahren. Man kann den Riemannschen Raumbegriff betrachten als die Aufhebung des Vorurteils, daß die Krümmungsverhältnisse an einer Stelle des Raumes verbindlich sein müßten für die Krümmung an allen anderen. Um dieser Aussage Riemanns einen Sinn zu geben, war zunächst die Annahme notwendig, daß der Maßstab, welcher an jeder Stelle zur Bestimmung der Koeffizienten $g_{ik}$ der metrischen Fundamentalform $ds^2 = g_{ik} dx^i dx^k$ zur Anwendung gelangt, ein „starrer“ Maßstab sei.

Demgegenüber macht Weyl mit Recht geltend, daß die Annahme eines solchen starren Maßstabes einer radikalen Nahgeometrie zuwider sei, daß nur die Verhältnisse der $g_{ik}$ an einer Stelle, nicht ihre Absolutbeträge, sinngemäß festgelegt werden können, und dementsprechend setzt er für die Änderung $dl$ eines Eichmaßstabes von der Länge $l$ bei einer infinitesimalen Verschiebung $dx^i$ an:

$$ dl = l \varphi_i dx^i, \quad (1) $$

wobei die Proportionalitätsfaktoren $\varphi_i$ Funktionen des Ortes sind, Charakteristika der Maßverhältnisse des Raumes — ähnlich den $g_{ik}$.

Oder, wenn man (1) integriert:

$$ l = l_0 e^{\int \tau_i dx^i} \quad (2) $$


52
$(l_0 = l$ am Anfang der Verschiebung). Das Eichmaß ist im allgemeinen vom Wege abhängig (nicht integrabel), es sei denn, daß die Größen

$$f_{ik} = \frac{\partial \varphi_i}{\partial x^k} - \frac{\partial \varphi_k}{\partial x^i}$$

(3)

verschwinden. Über diese Größen $f_{ik}$ kann man laut ihrer Definition (3) die Identität aussprechen (die Dimensionenzahl der Mannigfaltigkeit sei 4):

$$\frac{\partial f_{ik}}{\partial x^i} + \frac{\partial f_{kl}}{\partial x^l} + \frac{\partial f_{il}}{\partial x^k} = 0 \quad i \neq k \neq l, \quad i, k, l = 1, 2, 3, 4.$$ (4)

Die formale Übereinstimmung dieser vier Gleichungen mit dem einen System der Maxwell'schen Gleichungen

$$\text{rot} \mathcal{E} + \frac{1}{c} \mathcal{H} = 0,$$

$$\text{div} \mathcal{H} = 0,$$

sowie einige weitere formale Analogien haben Weyl zu dem Schluß geführt, die $\varphi_i$ seien bis auf einen konstanten Proportionalitätsfaktor zu identifizieren mit den Komponenten $\Phi_i$ des elektromagnetischen Viererpotentials, die $f_{ik}$ entsprechend mit den elektromagnetischen Feldstärken $\mathcal{E}, \mathcal{H}$. In folgerichtiger Ergänzung der geometrischen Deutung der Gravitation durch die variablen Krümmungen des Riemannschen Raumes dachte sich Weyl den noch übrigbleibenden Teil physikalischer Wirkungen, das elektromagnetische Feld, ebenfalls als eine Eigenschaft der Maßverhältnisse des Raumes, charakterisiert durch die Variabilität des Eichmaßes. Es ist also zu schreiben:

$$l = l_0 e^{\int_j t d x}, \quad (\alpha = \text{Proportionalitätsfaktor}).$$ (2a)

Quantenmechanische Deutung der Theorie von Weyl.

Maße (2 a), welches Weyl im magnetischen Felde annimmt. Es bedurfte wohl einer ungewöhnlich klaren metaphysischen Überzeugung, die Weyl solchen elementarsten Erfahrungen zum Trotz nicht von dem Gedanken abgehen ließ, daß die Natur von diesen schönen ihr gebotenen geometrischen Möglichkeiten Gebrauch machen müsse. Er hielt an seiner Auffassung fest und entzog die eben geschilderten Widersprüche der Diskussion durch eine etwas dunkle Umdeutung des Begriffs „realer Maßangabe“, womit nun allerdings seiner Theorie ihr so prägnanter physikalischer Sinn genommen war und sie dadurch sehr an Überzeugungskraft verlor.

Auf diese abstrakte Ausgestaltung der Theorie brauche ich nicht einzugehen. Ich werde vielmehr zeigen, daß gerade der prägnanten ursprünglichen Fassung der Weylschen Theorie eine noch viel größere Spannkraft innewohnt, als ihr Urheber bereits wirksam gemacht hat, daß man nämlich in ihr nichts geringeres als einen folgernichtigen Weg zur Undulationsmechanik zu erblicken hat, unter deren Gesichtspunkten sie erst eine unmittelbar verständliche physikalische Bedeutung gewinnt.

Kapitel II. Die Undulationsmechanik von de Broglie und die Theorie von Weyl.

Als „Theorie von de Broglie“ bezeichne ich jene noch unvollkommene Vorstufe der Undulationsmechanik, in welcher die Wellenfunktion der Bewegung eines Elektrons (auf welche wir uns hier beschränken)

\[ \psi = e^{\frac{2 \pi i}{\hbar} W(x_i)} \quad i = 1, 2, 3, 4 \]  

aus einer vollständigen Lösung \( W \) der Hamilton-Jacobischen partiellen Differentialgleichung

\[ \left( \frac{\partial W}{\partial x^i} - \frac{c}{\hbar} \Phi_i \right) \left( \frac{\partial W}{\partial x^i} - \frac{c}{\hbar} \Phi^i \right) = - m_0^2 c^2 \]  

hervorgeht, wobei die Integrationskonstanten in bekannter Weise so zu bestimmen sind, daß \( \psi \) eine eindeutige Funktion des Raumes, d. h. \( W \) additivperiodisch wird, mit einem ganzzahligen Multiplum der Planckschen Konstanten als Periode.

Wenn man Ernst macht mit der radikalen Kontinuumsauffassung der Materie, mit der Auflösung des diskontinuierlich abgegrenzten Elektrons in eine stetig in Raum und Zeit veränderliche Feldgröße, wie es
durch diese de Brogliesche konsequenter durch die später zu betrachtende Schrödingersche Theorie nahegelegt \(^1\) wird, so gelangt man in eine außerordentliche prinzipielle Schwierigkeit, wenn man untersucht, welchen Sinn man überhaupt metrischen Aussagen innerhalb des Undulotionskontinuums beizulegen hat. Denn in diesem schwingenden und fluktuiierenden unendlich ausgebreiteten Medium, welches an die Stelle des abgegrenzten Elektrons getreten ist, findet man keine unveränderlichen Diskontinuitäten, keine starren Körper, welche als reproduzierbare Maßstäbe die Festlegung einer Maßbestimmung gestatten könnten.

Ich vertrete durchaus nicht die Auffassung, daß, um von Geometrie im atomaren Gebiete zu reden, eine ausführbare Meßvorschrift angegeben werden müsse; von einer solchen kann ja auch in der Elektronentheorie nicht die Rede sein. Aber wenn man irgend einen definierten Sinn mit einer metrischen Angabe verbinden will, scheint mir, ist das mindeste, was man verlangen kann: die Angabe irgend eines realen Gegenstandes (als „Prototyp“), auf welchen die metrische Aussage bereits bezogen ist: Eines Elektronendurchmessers oder -abstands usw., wenngleich eine solche Aussage noch in einem sehr problematischen Zusammenhang zu einer ausführbaren Messung stehen mag.


---

\(^1\) Es sprechen bekanntlich wichtige Gründe dafür, auf welche vor allem von Born und seinen Mitarbeitern hingewiesen wurde, daß der ganze Undulationsformalismus statistisch umzudeuten ist. Insofern die Ladungsichte als eine statistische Gewichtsfunktion umgedeutet wird, ist es unschwer einzusehen, daß dieselbe Unbestimmtheit hinsichtlich der Anwendbarkeit des Satzes der Identität, auf die wir hier hinweisen, sich hinüberträgt. Aber da jene Auffassung zunächst jede Interpretation in Raum und Zeit ablehnt, hat für die Beziehung zur Weylischen Raumlehre geringes Interesse.
Quantenmechanische Deutung der Theorie von Weyl.

auf den allgemeinen Weylschen Raumgriff zurückzuziehen und zu versuchen, ihn auf das Schrödingersche Kontinuum anzuwenden. Da enthält sich nun ein einfacher Zusammenhang.

§ 1. Nehmen wir einmal an, wir besäßen bereits einen Maßstab \( l \) der sich nach der Weylschen Vorschrift (2\( a \)) verändert, und führen ihn im \( \psi \)-Felde herum. Und zwar werde er mit der Strömungsgeschwindigkeit der Materie, der Gruppenvierungsgeschwindigkeit

\[
\dot{u} = \frac{dx^i}{d\tau} = \frac{1}{m_0} \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi \right) \quad .
\]

(7)
geführt.

Ich behaupte, mit dieser naheliegenden Vorschrift über den Weg wird Weyls Skalar \( l \) numerisch identisch mit dem de Broglieschen Feldskalar \( \psi \). Hierzu sind noch zwei Präzisierungen zu treffen:

In dem Weylschen Eichmaß war noch ein Faktor \( \alpha \) unbestimmt gelassen: für diesen mache ich die Hypothese, er sei gleich \( \frac{2\pi i}{\hbar c} \). Also

\[
l = l_0 \frac{2\pi i}{c} \int \frac{e}{c} \Phi \text dt^i .
\]

(2\( a \))

Schließlich noch: ich benutze nicht genau das \( \psi \) aus Gleichung (5), sondern das mit dem Faktor \( e^\frac{2\pi i}{\hbar c^2} \) versehene fünfdimensionale \( \psi \), wie es den Vorschlägen von Klein, Fock und Kudar entspricht, wobei unter \( \tau \) die Eigenzeit\(^{1)} \) zu verstehen ist. Es sei also jetzt

\[
\psi = e^\frac{2\pi i}{\hbar c} (W + m_0 c^2 \tau) \quad .
\]

oder

\[
\psi = e^\frac{2\pi i}{\hbar c} \left\{ \int \frac{\partial W}{\partial x^i} \text dx^i + m_0 c^2 \tau \right\} .
\]

Diese Größe \( \psi \) ist zu vergleichen mit dem entlang der Strömung des Kontinuums geführten Weylschen Eichmaße (2\( a \)). Man erhält:

\[
\frac{\psi}{l} = \frac{1}{l_0} \frac{2\pi i}{c} \frac{\hbar}{e} \left\{ \int \left( \frac{\partial W}{\partial x^i} - \frac{e}{c} \Phi \right) \text dx^i + m_0 c^2 \tau \right\} ,
\]

hier sind die \( dx^i \) gemäß der durch (7) angegebenen Strömung zu führen:

\[
\frac{1}{l_0} \frac{2\pi i}{c} \frac{\hbar}{e} \left\{ \int \left( \frac{\partial W}{\partial x^i} - \frac{e}{c} \Phi \right) \left( \frac{\partial W}{\partial x^i} - \frac{e}{c} \Phi \right) \text dx^i - m_0 c^2 \tau \right\} .
\]

\(^{1)} \) Diese Auffassung von \( \tau \), die auf Kudar, Ann. d. Phys. 81, 632, 1926, zurückgeht, steht durchaus in Übereinstimmung mit der kürzlich diskutierten Deutung als Winkelkoordinate der Eigenrotationsbewegung des Elektrons (Naturwissenschaften 15, 15, 1927). Denn dieser Drehwinkel ist als eine vom Elektron mitgeführte Uhr anzusehen. Er transformiert sich wie die Eigenzeit.
erfüllen. Es ist vielmehr zu schreiben
\[
\frac{dx_k}{dx_0} = \frac{u_k}{c} = \frac{\psi \bar{\psi}}{Q} \cdot \frac{e}{m_0 c} \left( \frac{\partial W}{\partial x} - \frac{e}{c} \Phi_k \right), \tag{7 a}
\]
wobei der Faktor
\[
Q = e \psi \bar{\psi} \sqrt{1 + \left( \frac{\hbar}{2 \pi i} \right)^2 \frac{\Box}{m_0^2 c^2} \frac{|\psi|}{|\psi|}} = e \psi \bar{\psi} \left( 1 - \frac{e}{m_0 c^2} \Phi_0 \right) \tag{14}
\]
als "Ruhladungsdichte" abgetrennt ist.

In dieser Bezeichnung erhält man
\[
e \Phi_0 = m_0 c^2 \left( 1 - \frac{Q}{e \psi \bar{\psi}} \right) \tag{11 a}
\]
und die erste Schrödingersche Gleichung lautet in fünfdimensionaler Fassung¹):
\[
\sum_i \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) = 0. \tag{10 a}
\]

Wir vergleichen jetzt die Strecke \( l \) (13) entlang der Strömung (7 a) mit dem Schrödingerschen Skalar \( \psi \). Man erhält für \( \psi/l \)
\[
\psi = \frac{\psi}{l} = \frac{\psi}{l} e^{\frac{2\pi i}{\hbar}} \int \sum_i \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \, dx_i,
\tag{7 a}
\]
(7 a) ergibt:

\[
\psi = \frac{\psi}{l} e^{\frac{2\pi i}{\hbar}} \int \sum_i \frac{\psi \bar{\psi}}{\psi \bar{\psi}} \frac{e}{mc} \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \, dx_i + \left( \frac{\partial W}{\partial x_5} - \frac{e}{c} \Phi_3 \right) \, dx_5,
\tag{11 a}
\]
(11 a) ergibt:

\[
\psi = \frac{\psi}{l} e^{\frac{2\pi i}{\hbar}} \int \sum_i \frac{\psi \bar{\psi}}{\psi \bar{\psi}} \frac{e}{mc} \sum_i \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \, dx_i.
\]

(8 a) ergibt:

\[
\psi = \frac{\psi}{l} e^{\frac{2\pi i}{\hbar}} \int \sum_i \frac{\psi \bar{\psi}}{\psi \bar{\psi}} \frac{e}{mc} \sum_i \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \, dx_i.
\]

Letzteres wegen (10 a). Man erhält also zunächst nicht \( \psi/l = \text{konst} \), sondern
\[
\psi = \frac{\psi}{l} e^{\frac{2\pi i}{\hbar}} \int \sum_i \frac{\psi \bar{\psi}}{\psi \bar{\psi}} \frac{e}{mc} \sum_i \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \left( \frac{\partial W}{\partial x_i} - \frac{e}{c} \Phi_i \right) \, dx_i.
\]

¹) Hierbei ist zu beachten, daß \( \Phi_0 \) seinerseits noch selbst eine erst zu bestimmende Unbekannte ist. Bekanntlich ist es ein noch unverstandenes Wunder, warum das gleiche nicht für die Potentiale \( \Phi_1, \Phi_2, \Phi_3, \Phi_4 \) gilt, wie man erwarten müßte. (E. Schrödinger, Ann. d. Phys. 82, 263, 1927.)

\( \Phi_3 \) ist \( m_0 c \) [vgl. (5 a)].

\[57\]
welches eine eindeutige Ortsfunktion ist\(^1\). Aber die Potentiale \(\Phi_k\) sind nur bis auf einen additiven Gradienten physikalisch festgelegt; führe ich statt ihrer

\[
\Phi_k^* = \Phi_k - \frac{\hbar c}{2 \pi i e} \frac{\partial}{\partial x_k} \ln |\psi|
\]

als Potentiale ein, was die elektromagnetischen Feldstärken unberührt läßt, so folgt \(\psi/l = \text{konst.}\)

Die auf der Resonanz der Wellen beruhende Eindeutigkeit des mit der Strömung mitgeführten Eichmaßes überträgt sich natürlich jetzt ohne weiteres aus der de Broglieschen auf die Schrödingersche Theorie, so daß wir den Überlegungen des 2. Kapitels hier nichts hinzuzufügen haben.


\(^1\) Man kann diese Beweisführung im Sinne der fünfdimensionalen Geometrie sinngemäß folgendermaßen aussprechen:

\[
\left( \frac{\lambda}{\lambda} x^i - c \Phi_i \right) \text{ ist parallel dem Fünferstrom } j_i = \sum_{\text{j}} \epsilon_{\text{j}i} j_{\text{j}i} \left( \frac{\lambda}{\lambda} x^i - c \Phi_i \right).
\]

d.\(x^i\) soll parallel dem Fünferstrom \(j^i\) gewählt werden.

Der Fünferstrom ist orthogonal auf sich selbst \(\sum_{i=1}^{5} j_{i} \cdot i_{i} = 0\); also ist \(j_{i}\)

\[
\text{auch orthogonal auf } d x^i \text{ und also } \sum_{i=1}^{5} \left( \frac{\lambda}{\lambda} x^i - c \Phi_i \right) d x^i = 0.
\]

Ich verbeuge diese schöne Formulierung einer Mitteilung von Herrn A. Landé. Hierbei ist die 5. Komponente des Fünferstroms \(j_{5i}\), \(\epsilon_{5i}\).
Elektron und Gravitation. I.

Von Hermann Weyl in Princeton, N. J.

(Eingegangen am 8. Mai 1929).


Einleitung.


Um zweier Gründe willen verspricht die Adaption der Pauli-Diracschen Theorie des spinnenden Elektrons an die allgemeine Relativität zu physikalisch fruchtbaren Ergebnissen zu führen. 1. Die Diracsche Theorie, in welcher das Wellenfeld des Elektrons durch ein Potential ψ mit vier Komponenten beschrieben wird, gibt doppelt zu viel Energieniveaus: man sollte darum, ohne die relativistische Invarianz preiszugeben, zu den zwei Komponenten der Paulischen Theorie zurück-
Hermann Weyl, Elektron und Gravitation. 1.

kehren können. Daran hindert das die Masse $m$ des Elektrons als Faktor enthaltende Glied der Diracschen Wirkungsgröße. Masse ist aber ein Gravitationseffekt; es besteht so die Hoffnung, für dieses Glied in der Gravitationstheorie einen Ersatz zu finden, der die gewünschte Korrektur herbeiführt. 2. Die Diracschen Feldgleichungen für $\psi$ zusammen mit den Maxwellschen Gleichungen für die vier Potentiale $f_p$ des elektromagnetischen Feldes haben eine Invarianzeigenschaft, die in formaler Hinsicht derjenigen gleicht, die ich in meiner Theorie von Gravitation und Elektrizität vom Jahre 1918 als Eichinvarianz bezeichnet hatte: die Gleichungen bleiben ungeändert, wenn man gleichzeitig

$$\psi \text{ durch } e^{i\lambda} \psi \text{ und } f_p \text{ durch } f_p - \frac{\partial \lambda}{\partial x_p}$$

ersetzt, unter $\lambda$ eine willkürliche Ortsfunktion in der vierdimensionalen Welt verstanden. Dabei ist in $f_p$ der Faktor $\frac{e}{c \hbar}$ aufgenommen (— $e$ Ladung des Elektrons, $c$ Lichtgeschwindigkeit, $\frac{\hbar}{2\pi}$ Wirkungsquantum). Auch die Beziehung dieser „Eichinvarianz“ zum Erhaltungssatz der Elektrizität bleibt unangetastet. Es ist aber ein wesentlicher und für den Anschluß an die Erfahrung bedeutungsvoller Unterschied, daß der Exponent des Faktors, den $\psi$ annimmt, nicht reell, sondern rein imaginär ist. $\psi$ übernimmt jetzt die Rolle, welche in jener alten Theorie das Einsteinsche $ds$ spielte. Es scheint mir darum dieses nicht aus der Spekulation, sondern aus der Erfahrung stammende neue Prinzip der Eichinvarianz zwingend darauf hinzuweisen, daß das elektrische Feld ein notwendiges Begleitphänomen nicht des Gravitationsfeldes, sondern des materiellen, durch $\psi$ dargestellten Wellenfeldes ist. Da die Eichinvarianz eine willkürliche Funktion $\lambda$ einschließt, hat sie den Charakter „allgemeiner“ Relativität und kann natürlich nur in ihrem Rahmen verstanden werden.

An den Fernparallelismus vermag ich aus mehreren Gründen nicht zu glauben. Erstens sträubt sich mein mathematisches Gefühl a priori dagegen, eine so künstliche Geometrie zu akzeptieren; es fällt mir schwer, die Macht zu begreifen, welche die lokalen Achsenkreuze in den verschiedenen Weltpunkten in ihrer verdrehten Lage zu starrer Gebundenheit aneinander hat einfrieren lassen. Es kommen, wie ich glaube, zwei gewichtige physikalische Gründe hinzu. Gerade dadurch, daß man den Zusammenhang zwischen den lokalen Achsenkreuzen löst, verwandelt sich der Eichfaktor $e^{i\lambda}$, der in der Größe $\psi$ willkürlich bleibt, notwendig
aus einer Konstante in eine willkürliche Ortsfunktion; d. h. nur durch diese Lockerung wird die tatsächlich bestehende Eichinvarianz verständlich. Und zweitens ist die Möglichkeit, die Achsenkreuze an verschiedenen Stellen unabhängig voneinander zu drehen, wie wir im folgenden sehen werden, gleichbedeutend mit der Symmetrie des Energieimpulsensors oder mit der Gültigkeit des Erhaltungssatzes für das Impulsmoment.

Bei jedem Versuch zur Aufstellung der quantentheoretischen Feldgleichungen muß man im Auge haben, daß diese nicht direkt mit der Erfahrung verglichen werden können, sondern erst nach ihrer Quantisierung die Unterlage liefern für die statistischen Aussagen über das Verhalten der materiellen Teilchen und Lichtquanten. Die Dirac-Maxwellsche Theorie in ihrer bisherigen Form enthält nur die elektromagnetischen Potentiale $f_p$ und das Wellenfeld $\psi$ des Elektrons. Zweifellos muß das Wellenfeld $\psi'$ des Protons hinzugefügt werden. Und zwar werden in den Feldgleichungen $\psi$, $\psi'$ und $f_p$ Funktionen derselben vier Raum-Zeitkoordinaten sein, man wird vor der Quantisierung nicht etwa verlangen dürfen, daß $\psi$ Funktion eines Weltpunktes $(t, xy\varepsilon)$ und $\psi'$ Funktion eines davon unabhängigen Weltpunktes $(t', x'y'\varepsilon')$ ist. Es ist naheliegend, zu erwarten, daß von den beiden Komponentenpaaren der Diracschen Größe das eine dem Elektron, das andere dem Proton zugehört. Ferner werden zwei Erhaltungssätze der Elektrizität auftreten müssen, die (nach der Quantisierung) besagen, daß die Anzahl der Elektronen wie der Protonen konstant bleibt. Ihnen wird eine zweifache, zwei willkürliche Funktionen involvierende Eichinvarianz entsprechen müssen.

Wir prüfen zunächst die Sachlage in der speziellen Relativitätstheorie daraufhin, ob und inwieweit bereits die formalen Erfordernisse der Gruppentheorie, noch ganz abgesehen von den mit der Erfahrung in Einklang zu bringenden dynamischen Differentialgleichungen, die Erhöhung der Komponentenzahl $\psi$ von zwei auf vier notwendig machen. Wir werden sehen, daß man mit zwei Komponenten auskommt, wenn die Symmetrie von links und rechts aufgehoben wird.

Zweikomponententheorie.

§ 1. Transformationsgesetz von $\psi$. Führt man im Raume mit den kartesischen Koordinaten $x, y, \varepsilon$ homogene projektive Koordinaten $x_0$ ein:

$$x = \frac{x_1}{x_0}, \quad y = \frac{x_2}{x_0}, \quad \varepsilon = \frac{x_3}{x_0},$$

61
so lautet die Gleichung der Einheitskugel

\[- x_0^2 + x_1^2 + x_2^2 + x_3^2 = 0. \quad (1)\]

Projiziert man sie vom Südpol auf die Aquatorebene \( z = 0 \), die als Träger der komplexen Variablen

\[ x + iy = \xi = \frac{\psi_2}{\psi_1} \]

betrachtet wird, so gelten die Gleichungen

\[
\begin{align*}
x_0 &= \overline{\psi}_1 \psi_1 + \overline{\psi}_2 \psi_2, \\
x_1 &= \overline{\psi}_1 \psi_3 + \overline{\psi}_2 \psi_1, \\
x_2 &= i(-\overline{\psi}_1 \psi_3 + \overline{\psi}_2 \psi_1), \\
x_3 &= \overline{\psi}_1 \psi_3 - \overline{\psi}_2 \psi_2.
\end{align*}
\]

(2)

\( x_a \) sind Hermitesche Formen von \( \psi_1, \psi_2 \). Die Variablen \( \psi_1, \psi_2 \) sowie die Koordinaten \( x_a \) kommen hier nur ihrem Verhältnis nach in Frage. Eine homogene lineare Transformation von \( \psi_1, \psi_2 \) (mit komplexen Koeffizienten) bewirkt eine lineare, reelle Transformation unter den Koordinaten \( x_a \): sie stellt eine Kollination dar, welche die Einheitskugel in sich überführt und auf der Einheitskugel den Drehinn ungeändert läßt. Es ist leicht zu zeigen und wohl bekannt, daß man auf diese Weise jede derartige Kollination einmal und nur einmal erhält.

Vom homogenen Standpunkt zum inhomogenen übergehend, fasse man jetzt \( x_a \) als Koordinaten in der vierdimensionalen Welt und (1) als die Gleichung des „Lichtkegels“ auf; und man beschränke sich auf solche lineare Transformationen \( U \) von \( \psi_1, \psi_2 \), deren Determinante den absoluten Betrag 1 hat. \( U \) bewirkt an den \( x_a \) eine Lorentztransformation, d. i. eine reelle homogene lineare Transformation, welche die Form

\[- x_0^2 + x_1^2 + x_2^2 + x_3^2\]

in sich überführt. Doch lehren die Formel für \( x_0 \) und unsere Bemerkung über die Erhaltung des Drehungssinnes auf der Kugel ohne weiteres, daß wir unter den Lorentztransformationen nur die ein einziges in sich abgeschlossenes Kontinuum bildenden \( A \) bekommen, welche 1. Vergangenheit und Zukunft nicht vertauschen und 2. die Determinante + 1, nicht −1, besitzen; diese freilich ohne Ausnahme. Durch \( A \) ist die lineare Transformation \( U \) der \( \psi \) nicht eindeutig festgelegt, sondern es bleibt ein willkürlicher konstanter Faktor \( e^{i\alpha} \) vom absoluten Betrag 1 zur Disposition. Man kann ihn normalisieren durch die Forderung, daß die Determinante von \( U \) gleich 1 sei, aber selbst dann bleibt eine Doppeldeutigkeit zurück. An der Einschränkung 1. möchte man festhalten; es ist eine der hoffnungsvollsten Seiten der \( \psi \)-Theorie, daß sie der Wesensverschiedenheit von Vergangenheit und Zukunft Rechnung tragen kann. Die Einschränkung 2. hebt die Gleichberechtigung von links und rechts auf.
können, muß man die Koordinaten so spezialisieren, daß die kongridente Drehung aller Achsenkreuze als eine orthogonale Transformation der Koordinaten erscheint. Dies ist sicher möglich, doch gehe ich hier darauf nicht näher ein.

§ 6. Elektrisches Feld. Wir kommen jetzt zu dem kritischen Teil der Theorie. Meiner Meinung nach liegt der Ursprung und die Notwendigkeit des elektromagnetischen Feldes in folgendem begründet. Die Komponenten $\psi_1$, $\psi_2$ sind in Wahrheit nicht eindeutig durch das Achsenkreuz bestimmt, sondern nur insoweit, daß sie noch mit einem beliebigen "Eichfaktor" $e^{i\lambda}$ vom absoluten Betrag 1 multipliziert werden können. Nur bis auf einen solchen Faktor ist die Transformation bestimmt, welche die $\psi$ unter dem Einfluß einer Drehung des Achsenkreuzes erleiden. In der speziellen Relativitätstheorie muß man diesen Eichfaktor als eine Konstante ansehen, weil wir hier ein einziges, nicht an einen Punkt gebundenes Achsenkreuz haben. Anders in der allgemeinen Relativitätstheorie: jeder Punkt hat sein eigenes Achsenkreuz und darum auch seinen eigenen willkürlichen Eichfaktor; dadurch, daß man die starre Bindung der Achsenkreuze in verschiedenen Punkten aufhebt, wird der Eichfaktor notwendig zu einer willkürlichen Ortsfunktion. Dann ist aber auch die infinitesimale lineare Transformation $dE$ der $\psi$, welche der infinitesimalen Drehung $d\Omega$ entspricht, nicht vollständig festgelegt, sondern $dE$ kann um ein beliebiges rein imaginäres Multiplum $i \cdot df$ der Einheitsmatrix vermehrt werden. Zur eindeutigen Festlegung des kovarianten Differentials $\delta \psi$ von $\psi$ hat man also außer der Metrik in der Umgebung des Punktes $P$ ein solches $df$ für jedes von $P$ ausgehende Linienelement $PP' = (dx)$ nötig. Damit $\delta \psi$ nach wie vor linear von $dx$ abhängt, muß

$$df = f_p (dx)^p$$

eine Linearform in den Komponenten des Linienelements sein. Ersetzt man $\psi$ durch $e^{i\lambda} \cdot \psi$, so muß man zugleich, wie aus der Formel für das kovariante Differential hervorgeht, $df$ ersetzen durch $df - d\lambda$.

Dies hat zur Folge, daß zur Wirkungsdichte $m$ das Glied

$$\frac{1}{\xi} f(\alpha) s(\alpha) = \frac{1}{\xi} f(\alpha) \cdot \psi^* S(\alpha) \psi = f_p \cdot \psi^* \Omega^p \psi$$  \hspace{1cm} (26)$$

zu addieren ist. $m$ bedeutet fortan die so ergänzte Wirkungsgröße. Es herrscht notwendig Eichinvarianz in dem Sinne, daß die Wirkungsgröße ungeändert bleibt, wenn

$\psi$ durch $e^{i\lambda} \cdot \psi$, $f_p$ durch $f_p - \frac{\partial \lambda}{\partial x_p}$
ersetzt wird, unter \( \lambda \) eine willkürliche Ortsfunktion verstanden. Genau in der durch (26) beschriebenen Weise wirkt nach der Erfahrung das elektromagnetische Potential auf die Materie. Wir sind daher berechtigt, die hier eingeführten Größen \( f_p \) mit den Komponenten jenes Potentials zu identifizieren. Der Beweis ist vollkommen, wenn wir zeigen, daß auch umgekehrt das \( f_p \)-Feld nach denselben Gesetzen von der Materie beeinflußt wird, welche nach der Erfahrung für das elektromagnetische Potentialfeld gelten.

\[
f_{pq} = \frac{\partial f_q}{\partial x_p} - \frac{\partial f_p}{\partial x_q}
\]

ist ein eichinvarianter schiefsymmetrischer Tensor und

\[
l = \frac{1}{4} f_{pq} \tilde{f}^{pq}
\]

(27)
die für die Maxwellsche Theorie charakteristische skalare Dichte. Der Ansatz

\[
h = m + a l
\]

(28)

\((a\) eine numerische Konstante) liefert durch Variation der \( f_p \) die Maxwellschen Gleichungen mit

\[
- \psi^p = - \psi^* \tilde{\psi}^p
\]

als der Dichte des elektrischen Viererstroms.

Die Eichinvarianz steht in engem Zusammenhang mit dem Erhaltungssatz für die Elektrizität. Weil \( h \) eichinvariant ist, muß \( \delta \int h \, d x \) identisch verschwinden, wenn bei festgehaltenen \( \psi^p(\alpha) \) die \( \psi \) und \( f_p \) gemäß

\[
\delta \psi = i \lambda \cdot \psi, \quad \delta f_p = - \frac{\partial \lambda}{\partial x_p}
\]

variert werden; \( \lambda \) ist eine willkürliche Ortsfunktion. Dies liefert eine identisch erfüllte Relation zwischen den materiellen und den elektromagnetischen Gleichungen. Wissen wir, daß die materiellen Gleichungen (im engeren Sinne) gelten, so folgt also daraus

\[
\delta \int h \, d x = 0,
\]

wenn nur die \( f_p \) gemäß der Gleichung \( \delta f_p = - \partial \lambda / \partial x_p \) variiert werden. Andererseits folgt aus den elektromagnetischen Gleichungen dasselbe für die infinitesimalen Variation \( \delta \psi = i \lambda \cdot \psi \) der \( \psi \) allein. Wenn \( h = m + a l \), erhält man beide Male

\[
\int \delta h \, d x = \pm \int \psi^* \tilde{\psi}^p \frac{\partial \lambda}{\partial x_p} \, d x = \pm \int \lambda \frac{\partial \psi^p}{\partial x_p} \, d x.
\]

Eine analoge Sachlage fanden wir vor beim Erhaltungssatz von Energieimpuls und Impulsmoment. Sie verknüpften die materiellen Gleichungen

64
ON THE THEORY OF CHARGED FIELDS

Report drafted and submitted by O. KLEIN.

I. — INTRODUCTION.

The discovery of the so called heavy electron or mesotron and the role it is supposed to play in nuclear physics for the occurrence of attractive forces at small distances — a role suggested by Yukawa (1) already before the discovery — would seem to mean a considerable enlargement of the region of applicability of the field concept, which has hitherto been limited by the self energy difficulties. In fact, while electrons could be treated unambiguously only down to distances large compared to the so called radius of the electron, the new particle would seem to present no principal difficulties before one approaches its own « radius », which, due to its larger mass, is about two orders of magnitude smaller than that of the electron. Moreover a characteristic length of just the order of magnitude of the electronic radius is introduced as a « Compton » wave length corresponding to the mass of the new particle, which, as shown by Yukawa, gives the range of the forces due to the corresponding field.

The logical consistency of this enlargement of the field concept would seem to require the removal of the self energy difficulty of the electron at least down to distances approaching the radius of the new particle. Considering the order of magnitude and range of nuclear

forces — the nuclear binding energies being comparable with the rest mass energy of the electron and the range of the forces of the order of the electronic radius — it would not seem unreasonable to assume that a theory explaining nuclear attractions would also account for the rest mass of the electron, the attractive forces required as a compensation of the Coulomb repulsion being of a similar nature as the nuclear forces. A necessary condition for such an explanation — which would mean a removal of at least two orders of magnitude of the point from where the higher frequencies are "cut off" — is that the new forces belonging to the heavy electron field are determined by means of the elementary electric charge in a similar way as the electromagnetic forces, so that no other independent constant than the mass of the new particle will appear in the theory. Here we shall not enter further on the interesting problem as to the uniqueness of such a field theory comprising also the charged fields corresponding to the new particle if the demands of the electronic self energy and those of general invariance are to be satisfied, but we shall show how a theory of this kind may be built up as a consequent generalization of the formal scheme of ordinary field theory. If this theory of the mesoton and its interaction with electromagnetic fields and ordinary particles prove correct, it ought to satisfy automatically the above demands as to the electronic self energy.

For the invariant formulation of the field theory the so-called five dimensional representation has proved useful as a very direct and simple means of expressing the fundamental conservation theorems for energy, momentum and electric charge and their relation to space-time translation invariance and gauge invariance. In the original form of that representation, which leads exactly to the Einstein-Maxwell theory of gravitation and electromagnetism, the field quantities are supposed not to contain the new auxiliary coordinate \( x^0 \). The fact that this coordinate appears as the canonical conjugate (apart from a constant factor) to the electric charge suggests, however, as a natural generalization the assumption that the field quantities contain also terms depending upon \( x^0 \) and representing charged fields, which resemble those represented by the solutions of the
quantum theoretical wave equations of electric particles (1). The existence of a smallest quantum of electricity would hereby require — as a classical model later to be replaced by a suitable quantization — a periodicity in \( x^0 \) corresponding to the length \( L_0 = \frac{\hbar e \sqrt{2x}}{c} \) where \( \hbar \) is Planck’s quantum of action, \( c \) the vacuum velocity of light, \( e \) the elementary quantum of electricity and \( x \) the Einstein gravitational constant (2). The period being invariant under a gauge transformation of \( x^0 \), this model will satisfy automatically the claim of gauge invariance.

The direct and general way it expresses the fundamental conservation and invariance theorems seems to make this representation a natural starting point for a general quantum field theory comprising also the charged fields, which are supposed to correspond to the mesotons. The field quantities, which may be taken in a form adapted to the demands of a generally relativistic Dirac wave equation in five dimensions (3), would moreover have to satisfy commutation relations corresponding to the Bose-Einstein statistics. While these quantities will represent particles with integer spin, the ordinary elementary particles would have to be represented by Dirac spinor wave functions. But just, as the field quantities contain parts depending upon \( x^0 \) representing charged fields and parts without \( x^0 \) representing the ordinary electromagnetic and gravitational fields, we shall assume that also the spinors consist of \( x^0 \)-free components representing neutral particles (neutron, neutrino) and of \( x^0 \) components representing charged particles (proton, electron). Probably the higher harmonics of the Fourier development with respect to \( x^0 \) which would correspond to multiply charged particles, have no physical significance. They may be avoided in a similar way as in the theory of electronic spin through the introduction of two-row

(2) Klein, Nature, l. c.

67
matrices. This description of charged and neutral particles would be consistent with the way protons and neutrons are treated in Heisenberg's theory of nuclear constitution, the matrix representation just mentioned corresponding to the so called isotopic spin. On the other hand this matrix representation is nearly connected with the gauge transformation.

The theory outlined, which may be derived from a variation principle, the Lagrangian containing as well the spinor as the tensor field quantities, will describe an interaction between the proton, the neutron, the electron and the neutrino through the intermediate of the charged and uncharged fields giving as it would seem a quantitative formulation of the considerations of Yukawa, Kemmer and Bhabha on nuclear fields of force (1). Especially the theory will contain no new physical constants other than the mass of the meson, and the nuclear interactions will, like the electromagnetic forces in the first approximation neglecting all direct gravitational effects depend upon $e^2$.

As to the rest mass of the new particle, which does not appear in the ordinary field equations, it might be introduced by the addition of a term to the Lagrangian without disturbing the invariance. But it is not impossible that a further development of the theory will make this somewhat arbitrary addition superfluous, the mass appearing as some sort of self energy determined by the other lengths entering in the theory.

As a lack of simplicity of the theory it should be mentioned that although the allowed transformations are limited to general four-dimensional co-ordinate transformations and gauge transformations it makes use of invariants and tensors belonging to a general five-dimensional Riemann space. In the older form of the theory, where the physical quantities do not depend upon $x^0$, this inconvenience has been avoided by means of a projective treatment (2). Even if it may


(2) Voir W. Pauli, Ann. de Phys., 1933, 18, 305 et 337.
seem doubtful whether this treatment could be extended to the generalized form of the theory developed below, this lack is probably more formal than real. As to the more or less arbitrary addition (in the field equations for the charged fields) and omission (in the Dirac equations) of mass terms connected with electric charges it may have to do with a real limitation of the theory, these terms being perhaps — as indicated above — related to different stages of the self-energy problem. Whether the circumstance that the unitary treatment of gravitational, electromagnetic and charged fields yielded by the theory does not also comprise the spinor particles, the Lagrangian of which is simply added to the common Lagrangian of the fields mentioned, must be regarded as a lack is perhaps doubtful. It should further be emphasized that although gravitational actions will be neglected in the following the structure of the theory is very essentially determined by the circumstance that it fulfils from the outset the claim of general relativity. This indirect effect of gravitation may perhaps be compared with the influence of spin on the constitution of atoms and molecules also in the approximation where all direct effects of the spin momenta are neglected.

II. — MATHEMATICAL TREATMENT OF THE SPINOR PARTICLES.

Turning now to the mathematical formulation of the above remarks we introduce a Riemann metric tensor $\gamma_{\nu\mu}$, $\mu$ and $\nu$ taking the values 0, 1, 2, 3, 4. Corresponding to the invariance towards general transformations of the space-time co-ordinates $x^1, x^2, x^3, x^4 = \text{ict}$ together with that towards the gauge transformation expressed by means of the transformation equation $x'^\nu = x^\nu + \text{arbitrary function of space-time co-ordinates } (A)$, we define this tensor through the following relations

$$\gamma_{\nu\nu} = 1, \quad \gamma_{\nu k} = \beta \gamma_k, \quad \gamma^{kl} = g^{kl}, \quad k, l = 1, 2, 3, 4 \quad (I)$$

Here the $\gamma_k$ are certain functions of $x^0, x^1, x^2, x^3, x^4$, to be more closely characterized below, $\beta$ a constant equal to $\sqrt{2\alpha}$ and $g^{kl}$ the 10 contravariant components of the Einstein metric tensor. Further
we introduce a set of generally relativistic Dirac matrices $\gamma_\mu$ satisfying the relations

$$\frac{1}{2} (\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu) = \gamma_\mu \nu$$

(2)

and the corresponding quantities $\gamma_\mu = \gamma_\mu \nu$, which fulfil the relations and

$$\frac{1}{2} (\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu) = \delta_\mu \nu$$

(3)

and

$$\frac{1}{2} (\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu) = \gamma_\mu \nu$$

The $\gamma_\mu$ and the $\gamma_\mu$ may be taken as linear combinations of 5 ordinary constant Dirac matrices $\epsilon_0$, $\epsilon_1$, $\epsilon_2$, $\epsilon_3$, $\epsilon_4$ satisfying the relations

$$\frac{1}{2} (\epsilon_\mu \epsilon_\nu + \epsilon_\nu \epsilon_\mu) = \delta_\mu \nu$$

(4)

and since $\gamma_{00} = I$, we may put

$$\gamma_0 = \epsilon_0$$

(5)

Further it follows from the definition of the $\gamma_\mu$ that

$$\gamma_0 = \gamma_{00} \gamma_0 = \gamma_0 + \beta \chi_k \gamma^k$$

or

$$\gamma_0 = \epsilon_0 - \beta \chi_k \gamma^k$$

(6)

where the summation with respect to $k$ goes from 1 to 4.

We shall now consider the following expression

$$\phi \gamma^\rho \phi_\rho = \phi \gamma^\rho \gamma^\nu \phi$$

(7)

well-known from the Lagrangian leading to the generally relativistic form of Dirac's wave equations, the summation with respect to $\rho$ being here extended from 0 to 4, where

$$\phi_\rho = \frac{\partial \phi}{\partial x^\rho} - i \gamma^\rho \phi, \quad \phi_i = \frac{\partial \phi}{\partial x^i} + \phi \Gamma_i$$

(8)

are so called covariant derivatives of $\phi$ and $\phi$ respectively, the $\Gamma_i$ being certain matrices, the necessity of which was first shown by Schrödin-
ger. Introducing the expression (6) for $\gamma^i$ and omitting the term
\[ \varphi \frac{\partial \psi}{\partial x^0} - \frac{\partial \varphi}{\partial x^0} \epsilon_0 \psi \] — which omission does not violate the invariance required although from a formal point of view it is not altogether satisfactory — we get
\[ \varphi \gamma^k (\nabla_k \varphi) - (\nabla_k \gamma^i) \gamma^{i \varphi} - \varphi (\gamma^2 \Gamma^i + \Gamma^i \gamma^2) \]  
where
\[ \nabla_k = \frac{\partial}{\partial x^k} - \beta x_k \frac{\partial}{\partial x^0} \]  
(7 a)

From the assumption made above (and to be stated more precisely below) about the $x^0$ dependance of the quantities concerned, it follows that $\beta \frac{\partial}{\partial x^0}$ will not contain and will have to be retained when—as now we shall do all terms in the expression (7a) containing $\beta$ or any power of it are omitted, which means that we neglect all direct gravitational effects. Using a cartesian co-ordinate system ($g_{kl} = \delta_{kl}$) only $\Gamma_0$ will be «finite» with this approximation and given by
\[ \Gamma_0 = \frac{1}{2} \beta \frac{\partial x_k}{\partial x^0} \epsilon_0 \epsilon_k \]  
but the term $\gamma^0 \Gamma_0 + \Gamma_0 \gamma^0$ will be small of the order $\beta$ since $\epsilon_0$ anticommutes with all four $\epsilon_k$, and we are left with the expression
\[ \varphi \gamma^k (\Delta_k \varphi) - (\Delta_k \gamma^i) \gamma^{i \varphi} \]  
(7 b)
only, which by a suitable choice of the $\gamma^k$ will also be correct in a non-cartesian co-ordinate system, even in the non-Euclidean case (1).

As Lagrangian $L^0$ for our spin particles we take now
\[ L^0 = -\hbar c [\varphi \gamma^k (\nabla_k \varphi) - (\nabla_k \gamma^i) \gamma^{i \varphi}] + 2M c^2 \varphi \]  
(10)
where the mass $M$ of the particle is to be taken very nearly equal to the proton mass for the pair proton-neutron and probably equal to $0$ for the pair electron-neutrino.

(1) By a suitable choice of the $\gamma^k$ we might have made $\gamma^2 \Gamma^i + \Gamma^i \gamma^2$ equal to 0 from the beginning, but such a choice would in general not correspond to $\gamma_0 = \Gamma_0$. 
In a Cartesian co-ordinate system we may put

$$\gamma^k = \epsilon_k$$  \hspace{1cm} (11)

and further

$$\varphi = \psi^* r_4$$  \hspace{1cm} (12)

$\psi^*$ being the conjugate complex wave function to $\varphi$.

The more general expression (10) may be used in a well-known way to derive the components $T^{0}_{\kappa\lambda}$ of the corresponding energy-momentum tensor, which — since $L^0$ will have the numerical value $c$ — may be defined as by means of the relation

$$\delta L^0 = - T^{0}_{k\lambda} \delta g^{k\lambda}$$

obtained by putting

$$\delta \gamma^k = \frac{1}{2} \gamma^l \delta g^{kl}$$  \hspace{1cm} (13)

This gives

$$T^0_{k\lambda} = \frac{\hbar c}{4} \left[ \gamma^l \left( \nabla_k \varphi \right) + \varphi \gamma^l \left( \nabla_k \varphi \right) - \left( \nabla_k \varphi \right) \gamma_l \varphi - \left( \nabla_l \varphi \right) \gamma_k \varphi \right]$$  \hspace{1cm} (14)

where in a Cartesian co-ordinate system

$$\gamma_k = \epsilon_k$$  \hspace{1cm} (11 :)

In conformity with what has been said above about the $x^0$-dependence of the quantities concerned we put now

$$\chi_k = (A_k, \bar{A}_k), \quad \beta \frac{\partial \chi_k}{\partial x^0} = \frac{i e}{\hbar c} \left( 0, - \bar{B}_k \right)$$  \hspace{1cm} (15)

$$\varphi = (\psi_x, \psi_p), \quad \beta \frac{\partial \varphi}{\partial x^0} = \frac{i e}{\hbar c} \left( \psi_x, 0 \right)$$  \hspace{1cm} (15 a)

$$\varphi = (\varphi_x, \varphi_p), \quad \beta \frac{\partial \varphi}{\partial x^0} = - \frac{i e}{\hbar c} \left( 0, \varphi_p \right).$$

$\chi_1, \chi_2, \chi_3, \varphi$ being Hermitian matrices and $\chi_4$ a Hermitian matrix multiplied by $i$. The $A_k$ will be seen to be the components of the electromagnetic potential four-vector, while the $B_k$ and $\bar{B}_k$ describe the charged field. After a simple calculation we get from (10) and (13)
NEW THEORIES IN PHYSICS

\[ L^o = -\hbar c \left[ \begin{array}{c}
\varphi_N \gamma^k \frac{\partial \psi_N}{\partial x^k} - \frac{\partial \varphi_N}{\partial x^k} \gamma^k \psi_N + \varphi_P \gamma^k \frac{\partial \psi_P}{\partial x^k} - \frac{\partial \varphi_P}{\partial x^k} \gamma^k \psi_P - 2 \frac{i e}{\hbar c} A_k \varphi_P \gamma^k \psi_P - \\
\hbar c \left( \frac{B_k \varphi_N \gamma^k \psi_P + B_k \varphi_P \gamma^k \psi_N}{4} \right) + 2Mc^2 (\varphi_N \psi_N + \varphi_P \psi_P) \right] \]  

(16)

\[ T_{k1}^o = \frac{\hbar c}{4} \left[ \begin{array}{c}
\varphi_N \gamma_k \frac{\partial \psi_N}{\partial x^l} + \varphi_N \gamma^l \frac{\partial \psi_N}{\partial x^k} - \frac{\partial \varphi_N}{\partial x^l} \gamma_k \psi_N + \\
\varphi_P \gamma_k \frac{\partial \psi_P}{\partial x^l} + \varphi_P \gamma^l \frac{\partial \psi_P}{\partial x^k} - \frac{\partial \varphi_P}{\partial x^l} \gamma_k \psi_P - \\
2 \frac{i e}{\hbar c} (A_k \varphi_P \gamma^k \psi_P + A_k \varphi_P \gamma^k \psi_P) - \\
\frac{i e}{\hbar c} (B_k \varphi_N \gamma_1 \psi_P + B_k \varphi_N \gamma_1 \psi_P + B_k \varphi_P \gamma_1 \psi_N + B_k \varphi_P \gamma_1 \psi_N) \right] \]  

(17)

The expression (16) for \( L^o \) may further be used to determine the current-density four-vector \( S_{0k} \) since

\[ S_{0k} = \frac{i e}{2} \frac{\partial L^o}{\partial A_k} \]  

(18)

In this way we get

\[ S_{0k} = i e \varphi_P \gamma^k \psi_P \]  

(19)

In a Cartesian co-ordinate system this gives according to (12)

\[ S_{0k} = i e \varphi^* \epsilon_i \epsilon_k \psi_P \]  

(20)

and especially

\[ S^{ii} = i e \varphi^* \psi_P \]

in conformity with Dirac's original assumption about \( \psi^* \psi \). We see also that only \( \psi_P \) and not \( \psi_N \) contributes to the electric current and density.

From the expression (16) for \( L^o \) and by means of the variation principle

\[ \delta \int L^o \sqrt{g} \ dx = 0, \ g = \left| g_{k1} \right|, \ dx = dx^1 \ dx^2 \ dx^3 \ dx^4 \]  

(21)

the variations of \( \varphi \) and \( \psi \) vanishing at the border of the space-time region under consideration, we obtain the following wave equations

73
for $\psi_N, \bar{\psi}_N, \psi_P, \bar{\psi}_P$, where for the sake of simplicity we have used a Cartesian co-ordinate system

$$\varepsilon_k \left( \frac{\partial \psi_N}{\partial x^k} - ie \frac{B_k \psi_P}{2\hbar c} \right) = \frac{Mc}{\hbar} \psi_N \left( \frac{\partial \bar{\psi}_N}{\partial x^k} + \frac{ie}{2\hbar c} B_k \bar{\psi}_P \right) \varepsilon_k = - \frac{Mc}{\hbar} \bar{\psi}_N. \quad (22)$$

$$\varepsilon_k \left( \frac{\partial \psi_P}{\partial x^k} - ie \frac{A_k \psi_P}{\hbar c} - \frac{ie}{2\hbar c} B_k \bar{\psi}_N \right) = \frac{Mc}{\hbar} \psi_P \left( \frac{\partial \bar{\psi}_P}{\partial x^k} + \frac{ie}{\hbar c} A_k \bar{\psi}_N + \frac{ie}{2\hbar c} B_k \bar{\psi}_N \right) \varepsilon_k = - \frac{Mc}{\hbar} \bar{\psi}_P,$$

which are seen to be of the Dirac type, the terms containing the $B$-field, which do not appear in the ordinary equations, describing the action on the spinor particles due to the charged fields or heavy electrons.

III. — TREATMENT OF CHARGED AND NEUTRAL FIELDS.

In order to obtain field equations for the $A$-and $B$-fields we shall now regard the quantity $\Gamma$, which is the five-dimensional analogon to the Lagrangian of the Einstein gravitational field equations, namely

$$\Gamma = \gamma_{\lambda^\nu} \left[ \frac{\mu}{\sigma} \left( \right) \right] \left( \frac{\bar{\sigma}}{\tau} \right) \left( \right) \left( \frac{\bar{\tau}}{\tau} \right) - \frac{\mu}{\sigma} \left( \right) \left( \frac{\bar{\sigma}}{\tau} \right) \left( \right) \left( \frac{\bar{\tau}}{\tau} \right) \right] \quad (23)$$

$$\left( \begin{array}{c} \mu \\ \nu \\ \rho \end{array} \right) \right) = \frac{1}{2} \gamma_{\tau^\mu} \left( \frac{\partial \gamma_{\nu^\mu}}{\partial x^\nu} + \frac{\partial \gamma_{\nu^\tau}}{\partial x^\mu} \right) \left( \frac{\gamma_{\nu^\mu}}{\partial x^\tau} \right) \quad (24)$$

To begin with we consider only the case where gravitation may be altogether neglected using a Cartesian co-ordinate system, where the Einstein metric tensor is given by

$$g_{\lambda^\nu} = \delta_{\lambda^\nu} \quad (25)$$

and further we neglect all quantities containing higher powers of $\beta$ than the second. Using the expressions (1) we get after a simple calculation

$$\Gamma = - \frac{\beta^2}{4} \chi_{\nu^\mu} \chi_{\nu^\mu} \quad (26)$$

74
NEW THEORIES IN PHYSICS

where

\[ \chi_{rs} = \nabla_r \chi_s - \nabla_s \chi_r \]  \hspace{1cm} (27)

\( \nabla \), being the operator defined in (9).

In a non Cartesian co-ordinate system the corresponding formula will be (where we have still neglected those terms in \( g_{ki} \) which depend upon \( x^0 \))

\[ \Gamma = - \frac{\beta^2}{4} \ g^{kr} \ g^{is} \ \chi_{kl} \chi_{rs} + G \]  \hspace{1cm} (28)

where \( G \) is the Einstein Lagrangian formed by means of the \( g_{ki} \) in a similar way as (23) is formed by means of the \( \gamma_{\mu \nu} \). Apart from the terms containing \( x^0 \), where we have neglected all terms of gravitational order of magnitude, the formula (28) is exact and contains the well-known five-dimensional representation of the Maxwell-Einstein field theory, the \( \chi_{rs} \), being apart from \( x^0 \)-terms the components of the electromagnetic field six-vector. Although (28) has a generally relativistic form and therefore may be expected to give a correct description of the influence of ordinary gravitational fields it will probably not give a consequent treatment of the finer gravitational effects, since the neglected terms — among others these corresponding to charged \( g_{kl} \) — fields — are of the same gravitational order of magnitude. For our present purpose, however, which is to determine the energy-momentum tensor of the charged field in the non-gravitational case it will suffice.

Substituting now for the \( \chi_r \) the expression (15) we get

\[ \chi_{rs} = \begin{pmatrix} A_{is} & B_{is} \\ B_{rs} & A_{rs} \end{pmatrix} \]  \hspace{1cm} (29)

where

\[ A_{rs} = F_{rs} + \frac{ie}{\hbar c} (B_r \hat{B}_s - B_s \hat{B}_r) \text{ and } F_{rs} = \frac{\partial A_s}{\partial x^r} - \frac{\partial A_r}{\partial x^s} \]  \hspace{1cm} (30)

\( F_{rs} \), being the components of the ordinary electromagnetic field six-vector and
We have now to substitute the diagonal term of $\chi_{rs}$ $\chi_{rs}$, namely $A_{rs} A_{rs} + B_{rs} \dot{B}_{rs}$ into our Lagrangian (26) giving if we go over to the generally relativistic form

$$\Gamma = -\frac{x}{2} \delta_{kl} \delta_{is} (A_{kl} A_{rs} + B_{kl} \dot{B}_{rs}) + G,$$  \hspace{1cm} (32)

where we have replaced $\frac{\beta^2}{2}$ by $x$. To obtain the total Lagrangian we have further to add a term corresponding to the rest mass $\mu$ of the mesoton, namely $-\frac{\mu^2 c^2}{\hbar^2} g^{kl} B_k B_k$ and also the quantity $x L$, which gives

$$L = \Gamma + x L^0 = x \frac{\mu^2 c^2}{\hbar^2} g^{kl} B_k B_k$$  \hspace{1cm} (33)

In the variation principle

$$\delta \int L \sqrt{-g} \, dx = 0$$  \hspace{1cm} (34)

the quantities to be varied are now $\psi_N$, $\varphi_N$, $\psi_r$, $\varphi_r$, $A_k$, $B_k$, $\dot{B}_k$, and $g^{kl}$, $k, \ i = 1, 2, 3, 4$, all variations vanishing at the border of the region. The variation with respect to the $\psi : s$ and $\varphi : s$ give, as we have already seen, the equations (22). The variation with respect to $g^{kl}$ gives us the Einstein field equations with right sides containing the energy-momentum tensor, namely

$$G_{kl} = R_{kl} - \frac{i}{2} g_{kl} R = -\frac{1}{4} T_{kl}$$  \hspace{1cm} (35)

where

$$T_{kl} = T^0_{kl} + T^e_{kl}$$  \hspace{1cm} (36)

$T^0_{kl}$ being the quantities defined by (17) while

$$T^e_{kl} = A_{kr} A_{lr} + \frac{i}{2} (B_{kr} \dot{B}_{lr} + B_{lr} \dot{B}_{kr}) - \frac{i}{4} \delta_{kl} (A_{rs} A_{rs} + B_{rs} \dot{B}_{rs})$$  \hspace{1cm} (37)
R_{kl} are the components of the well-known Einstein curvature tensor and R the corresponding invariant. In the formula for T_\mu, the components of the energy-momentum tensor due to the A- and B-fields, we have used for the sake of simplicity a Cartesian co-ordinate system. If the terms containing the B : s are omitted, we obtain the usual formula for the electromagnetic energy-momentum tensor.

A variation of the A^\mu gives after a simple calculation, where again we use a Cartesian co-ordinate system,

\[
\frac{\delta T_\mu}{\delta x^k} = S_k = S_k^0 + S_k^\mu \tag{38}
\]

with

\[
S_k^\mu = \frac{ie}{\hbar c} \left[ \frac{\delta}{\delta x^\mu} (B_k B_\mu - B_\mu B_k) + \frac{i}{2} (B_{kl} B_k - B_{kl} B_k) \right] \tag{39}
\]

while

\[
S_k^0 = ie \phi_\nu \epsilon^\nu \phi_\mu \tag{19a}
\]

is the spinor particle current given above. We see that the equations (39) are the ordinary Maxwell equations for the electromagnetic field, the right side containing the current-density four-vector, to which the charged B-field contributes by the quantities s^\mu_k.

Finally we get wave equations for B_k and B_k by varying B_k and B_k respectively, namely

\[
\frac{\delta}{\delta x^i} B_{kl} + \frac{ie}{\hbar c} A_{kl} B_i + \frac{\mu^2 c^3}{\hbar^2} B_k = ie \phi_\nu \epsilon_k \phi_\mu \tag{40}
\]

We may easily verify the conservation theorem for the total electric charge. Indeed a simple calculation gives

\[
\frac{\delta S_k^\mu}{\delta x^k} = \frac{ie}{2\hbar c} \left[ B_i (\delta_k B_{kl}) - B_k (\delta_k B_{kl}) \right]
\]

or by means of (40)

\[
\frac{\delta S_k^\mu}{\delta x^k} = \frac{e^2}{2\hbar c} (B_k \phi_\nu \epsilon_k \phi_\mu - \dot{B}_k \phi_\nu \epsilon_k \phi_\mu). \tag{41}
\]
From (22) we get similarly
\[ \frac{\partial S^0_k}{\partial x^k} = -\frac{e^2}{2\hbar c} (B_k \varphi_p \varepsilon_k \psi_n - \hat{B}_k \varphi_n \varepsilon_k \psi_p) \quad (41) \]
showing that
\[ \frac{\partial S_k}{\partial x^k} = \frac{\partial S^0_k}{\partial x^k} + \frac{\partial S^e_k}{\partial x^k} = 0 \]
Thus the total charge is conserved, but not the charge of the spinor particles or that of the charged fields separately. At the same time the number of spinor particles (charged and uncharged) is conserved. In fact the sum of the probability current-density vectors belonging to the neutral and the charged spinor particles respectively, \( i \varphi_n \varepsilon_n \psi_n + i \varphi_p \varepsilon_k \psi_p \), fulfils the continuity equation, in that
\[ \frac{\partial}{\partial x^k} (i \varphi_n \varepsilon_k \psi_n) = \frac{e}{2\hbar c} (B_k \varphi_p \varepsilon_k \psi_n - \hat{B}_k \varphi_n \varepsilon_k \psi_p) \]

The equations (40) for the B-field are similar to though not identical with the equations given by Proca (1) and applied by several authors to the problem of the mesoton. The most important difference is the appearance of non-linear terms in (40) due to the quantities \( \frac{ie}{\hbar c} (B_r B_s - \hat{B}_r \hat{B}_s) \) in the \( A_{rs} \). These same quantities are also seen to appear in the expression for the current-density vector, where they give rise to a magnetic spin moment. In fact, the component of the magnetic moment due to that part of the current in the direction of \( x^1 \) is given by the following expression
\[ \frac{ie}{\hbar c} \int (\hat{B}_2 B_2 - \hat{B}_2 \hat{B}_2) \, dr = dx^1 \, dx^2 \, dx^3, \]
where the integral has to be taken over all space, the expressions \( \frac{ie}{\hbar c} (\hat{B}_k B^k - \hat{B}_k \hat{B}_k) \) defining thus the density of magnetic spin moment (2).

---

(2) See Proca, J. Phys. Radium, l.c.
It should perhaps be pointed out once more that the Lagrangian $L_0$ may belong either to the pair neutron-proton or to the pair neutrino-electron. In (3.3) $L_0$ should therefore, strictly speaking, be the sum of two different $L_0$:s, the one referring to the heavy and the other to the light spinor particles, each with its own set of $\psi$:s. This we have omitted for the sake of shortness and because the corresponding completion is quite obvious. But it is worth while to notice that the complete Lagrangian will imply an interaction of heavy and light spinor particles not only through the intermediate of the electromagnetic field but also through the B-field, an interaction which will entail the occurrence of $\beta$-processes, the probability of which may be calculated on the basis of the theory developed in this report.

This theory, however, will not be completely founded before we have given rules for its quantization, a question we shall briefly touch upon here. First of all the $\psi$ corresponding to the different kinds of spinor particles have to satisfy well-known quantization rules, being the application of the formalism of Jordan-Wigner to the wave equation of Dirac and corresponding to Fermi statistics. Moreover the electromagnetic quantities have to be quantized in the usual way, so that the remaining problem is the formulation of quantization rules for the B-field. Here we meet with the same formal difficulty as in the case of the electromagnetic field that the time derivatives of $B_i$ and $\dot{B}_i$ are not present in the Lagrangian. We may, however, get over this difficulty by means of the general method developed by Rosenfeld. As a result one obtains the following commutation relations (1)

\[
[B_{ik}(r), B_{kl}(r')] = 2\hbar c \delta_{kl} \delta(r-r')
\]

\[
[B_{ik}(r), \dot{B}_{kl}(r')] = 2\hbar c \delta_{kl} \delta(r-r')
\]

where $r$ and $r'$ indicate two space points to be taken at the same time $t$, $\delta(r-r')$ being the well-known singular function. Further the $B_k$ com-

(1) Similar relations for the Proca equations are given by Kemmer and Bhabha, Proc. Roy. Soc., l. e.
mute with all the $B_i$, $B_k$ and all the $B_{kl}$, the $B_k$ similarly with all the $B_k$, $B_k$ and all the $B_{kl}$, while $B_i$ and $B_i$ commute with all the quantities mentioned.

As an interesting application of the commutation rules we may inquire into the commutation expression for the total charge $Q^e$ of the B-field with one of the quantities $B_k$, $B_k$. We have

$$ Q^e = -i \int S_i^e (r') \, dr' $$

$S_i^e$ being given by (39). A simple calculation gives now

$$ [Q^e, B_k (r)] = \frac{e}{\hbar c} \int [ \dot{B}_{ik} (r'), B_k (r) ] B_k (r') \, dr' = e B_k (r) \quad (43) $$

and similarly

$$ [Q^e, \dot{B}_k (r)] = -\frac{e}{\hbar c} \int [ B_{ik} (r'), \dot{B}_k (r) ] \dot{B}_k (r) \, dr' = e \dot{B}_k (r) \quad (44) $$

Since

$$ \frac{\partial B_k}{\partial \xi^o} = \frac{ie}{\hbar c} B_k, \quad \frac{\partial \dot{B}_k}{\partial \xi^o} = -\frac{ie}{\hbar c} \dot{B}_k, $$

we see that $Q^e$ as far as the B-field is concerned plays the role of the canonical conjugate to $\xi^o$, which is a special case of a general rule according to which $\xi^o$ is canonically conjugate to the total charge divided by $\beta c$. At the same time the relations (43, 44) are seen to be closely related to the fact that the charge is always a whole positive or negative multiple of the elementary electric charge. They express moreover the gauge transformation properties of the $B_k$ and $\dot{B}_k$ (1).

** * * *

**PROFESSOR MOELLER** recalled that discussions of the experimental results relative to the interactions between the particles constituting

nuclei appeared to show the existence of forces between two protons and between two neutrons of the same order of magnitude as the force between one proton and one neutron. If we wished to explain these forces by the intermediacy of a field of particles capable of being given out or absorbed by protons and neutrons, we must then admit, in addition to Yukawa's charged particles, neuter particles of the same mass. It was hard to see how such neuter particles could naturally find a place in a general scheme of the type proposed by Professor Klein.

Professor Klein answered that if, in the Lagrangian, we replace the $\gamma_{rs}$ given by (29), by the following expression

$$\gamma_{rs} = \left( \begin{array}{c} A_{rs}' - C_{rs}, B_{rs} \\ B_{rs}, A_{rs} + C_{rs} \end{array} \right)$$ \hspace{1cm} (29a)$$

where

$$C_{rs} = \frac{\partial C_a}{\partial x} - \frac{\partial C_r}{\partial x}$$

and, in the Lagrangian (10), the $\gamma_r$ given by (15), by the corresponding expression,

$$\gamma_r = \left( \begin{array}{c} A_r - C_r, B_r \\ B_r, A_r + C_r \end{array} \right)$$ \hspace{1cm} (15a)$$

The C-field corresponds to neuter particles to which it is possible to give any mass, for instance that of the mesoton. It seems, however, that such particles would give rise, according to the Lagrangian concerned, to repressive forces and not to attractive forces between protons.
I. INTRODUCTION

These lectures are a review of the recent experimental results obtained at the CERN pp Collider ($\sqrt{s} = 540 \text{ GeV}$) on the study of hard collisions, defined as those leading to high transverse momentum ($p_T$) secondaries.

In the language of the parton model, the two colliding bodies are not the incident hadrons, but two of their point-like constituents (partons, i.e. quarks or gluons), each carrying a large fraction of the incident $p$ and $\bar{p}$ momenta. Thus the relevant value of the collision energy is not $\sqrt{s}$, but rather $x_1 x_2 s$, where $x_1$ and $x_2$ are the fractions of the incident beam momenta carried by the two partons.

This written version has been updated with respect to the lectures delivered in Tâbor to include most of the very important results which became available shortly after the end of the 1983 JINR-CERN School of Physics.

These lectures begin with a short historical survey on the physics of high-$p_T$ hadronic final states before the first Collider operation (Section 2). These final states are believed to result from elastic or quasi-elastic scattering of two incident partons, and they consist, in general, of collimated jets of high-$p_T$ hadrons. After a description of the CERN pp Collider facility and of the two major experiments being performed with it (Section 3), the spectacular results obtained at the Collider on the production of hadronic jets are discussed in detail in Section 4.

Finally, in Section 5 another class of hard collisions is discussed, namely quark-antiquark (qq) annihilation into a heavy state with mass $M = \sqrt{x_1 x_2 s}$. In this case high-$p_T$ secondaries result from the decay of the heavy state which has been produced. The most impressive example of such collisions is the production of the weak Intermediate Vector Bosons (IVBs) which have been postulated to mediate the electroweak interaction.

2. HIGH-$p_T$ HADRONIC FINAL STATES: HISTORICAL SURVEY

The interest in the possible existence of a kind of hadron collision leading to high-$p_T$ secondaries began at the end of the sixties. In order to interpret the SLAC results on electron-nucleon scattering in the deep inelastic region$^1$) Bjorken suggested in 1969 the scaling behaviour of the nucleon structure functions$^2$). In the same year Feynman$^3$) proposed the parton model, according to which a hadron, when described in a reference frame where its momentum is infinite, is composed of independent point-like constituents (partons).

These two ideas led Berman, Bjorken and Kogut$^4$) to predict the existence of a class of hadron collisions producing high-$p_T$ particles in the final state. According
to these authors a hard hadron collision can be described as a three-step process, as illustrated by the following example. Let us consider a $\bar{p}p$ collision at the CERN Collider. In the initial state a typical parton distribution in the $(p_L, p_T)$ plane is shown in Fig. 1a, where the sum of the longitudinal momenta $p_L$ of all partons belonging to the incident $p$ or $\bar{p}$ equals the beam momentum (±270 GeV/c in this case). We then assume that the partons with $p_L = +100$ and $-70$ GeV/c, respectively, undergo elastic scattering at an angle of $90^\circ$ in their centre-of-mass frame, resulting in the intermediate state shown in Fig. 1b. This state evolves by interacting with the other partons through a final-state interaction which will generally involve only low-$p_T$ mechanisms. The final result is shown in Fig. 1c, where the two high-$p_T$ partons have fragmented into hadrons having small transverse momentum with respect to the directions of the scattered partons. As also shown in Fig. 1c, these two hadronic jets are approximately coplanar with the beam axis, because the initial-state partons have low $p_T$.

![Diagram](image)

**Fig. 1** Representation of a hard hadron collision in momentum space as a three-step process.
It was, furthermore, predicted that the invariant cross-section for inclusive production of high-\( p_T \) hadrons at \( 90^\circ \) should have the form

\[
E \frac{d^3\sigma}{dp_T^3} = p_T^{-n} F(x_T),
\]

where \( F \) is a dimensionless function of the scaling variable \( x_T = 2p_T/\sqrt{s} \), and \( n = 4 \).

This prediction followed as a consequence of the scaling behaviour of both the structure functions of the incoming hadrons and the fragmentation functions of the outgoing high-\( p_T \) hadrons.

These expectations were confirmed by three ISR experiments\(^5-7\)), which found that the yield of charged and neutral pions as a function of \( p_T \) near \( 90^\circ \) exceeded a naive extrapolation from lower energy data by several orders of magnitude (this factor, which increased with \( p_T \), was \( \sim 10^7 \) at \( p_T = 6 \text{ GeV/c} \)). It was also observed that the inclusive cross-sections, when measured at fixed \( p_T \) as a function of \( \sqrt{s} \), showed a dependence upon \( \sqrt{s} \) which was stronger the higher the \( p_T \) value being considered.

This was in sharp contrast with the behaviour at low \( p_T \) (\( \leq 1 \text{ GeV/c} \)), where the inclusive cross-section increases only by \( \sim 20\% \) over the ISR energy range \( (22 < \sqrt{s} < 63 \text{ GeV}) \).

These observations are clearly illustrated by the data of Büscher et al.\(^7\)) on the reaction \( p + p \rightarrow \pi^0 + \text{anything} \), which are shown in Fig. 2. However, fitting these data to the functional form previously discussed [Eq. (1)] gave \( n = 8 \), in disagreement with a simple interpretation in terms of collisions between partons.

![Fig. 2 First observation of high-\( p_T \) \( \pi^0 \) mesons in proton-proton collisions. The invariant cross-section for the reaction \( p + p \rightarrow \pi^0 + \text{anything} \) is plotted as a function of \( p_T (\pi^0) \) for various values of \( \sqrt{s} \).](image-url)
FRIDAY - 10 June
9,00 - 10,15 L.B. Okun (4)
11,00 - 12,15 M.B. Voloshin (4)
15,30 - 17,00 DISCUSSION SESSIONS
17,30 - 18,45 F. Dydak (2)
20,00 H. Hopper:
THE CERN experimental programme

SATURDAY - 11 June
9,00 - 10,15 L.B. Okun (5)
11,00 - 12,15 M.B. Voloshin (5)
FREE
20,00 JINF-CERN BANQUET

SUNDAY - 12 June
ALL DAY EXCURSION

MONDAY - 13 June
9,00 - 10,15 F. Dydak (5)
11,00 - 12,15 C.W. Fabjan (6): EXPERIMENTAL METHODS, PHYSICS OF DETECTORS
15,15 - 17,00 DISCUSSION SESSIONS
17,30 - 18,45 F. Dydak
CFL, single-particle collider
20,00 CONCERT (city theatre)

TUESDAY - 14 June
9,00 - 10,15 F. Dydak (4)
11,00 - 12,15 G. Ross (1): GRAND UNIFICATION AND SUPERSYMMETRY
FREE + EXCURSION

WEDNESDAY - 15 June
9,00 - 10,15 G. Ross (1)
11,00 - 12,15 C.W. Fabjan (2)
15,30 - 17,00 DISCUSSION SESSIONS

JINR - CERN SCHOOL 87
In the following years much theoretical work was done in order to explain this result. Among the various models which were suggested, it is worth mentioning the so-called Constituent Interchange Model (CIM), which interpreted the observed $p_T$ dependence as due to the contributions from other subprocesses, such as the collision of a meson (considered here as a parton) and a quark. However, this model failed to explain the observed $\pi^-/\pi^+$ ratio for inclusive production at high $p_T$ in $\pi^-$-nucleon collisions at high energy and, in spite of other successes, it was ruled out.

On the experimental side the study of the structure of events containing high-$p_T$ particles in the final state represented one of the main research lines at the ISR and Fermilab. Most of the experiments used a trigger based on a single high-$p_T$ particle which distorted the structure of these events, and especially the structure of the jet to which the trigger particle belonged [this effect is known under the name of trigger bias]. However, in spite of this problem, most of the results were found to be consistent with the original predictions of the parton model.

In particular we quote:

- the approximate coplanarity between the trigger particle and the "opposite-side" jet, which is not distorted by the trigger bias;
- the limited transverse momentum of the high-$p_T$ hadrons with respect to the jet axis (its average value ($q_T$) was found to be $\sim$ 0.5 GeV/c);
- the approximate scaling behaviour of the fragmentation function of the "opposite-side" jet. If $P$ is the total jet momentum and $p$ is the momentum of a hadron in the jet, the fragmentation function $dn/dz$, where $z = p \cdot P/|p|^2$ is independent of $P$.

Further evidence in favour of the production of jets in hadron collisions came from the striking similarities between the characteristics of these jets (average multiplicity, $(q_T)$, fragmentation functions) and those of jets observed in $e^+e^-$ collisions and lepton-nucleon ($eN$ or $\nu N$) scattering in the deep inelastic region.

The apparent contradiction between the success of the parton model to explain the structure of high-$p_T$ events and the observation of a $p_T^{-3/4}$ scaling law instead of the expected $p_T^{-3}$ form was finally resolved in the late seventies with the advent of Quantum Chromodynamics (QCD). This theory introduced corrections to the parton model which explained the scaling violation effects observed in second-generation experiments on deep inelastic lepton-nucleon scattering and $e^+e^-$ annihilation into hadrons. When applied to hadron collisions, QCD corrections were found to break the simple scaling law given by Eq. (1), and the observed $p_T^{-3}$ scaling law was then interpreted as an approximate description of the data valid only for $p_T < 8$ GeV/c.

At the same time, a new series of ISR experiments extended the measurement of inclusive $\pi^0$ production up to $p_T \approx 16$ GeV/c and found that a constant expo-
nent \( n \) in Eq. (1) is no longer adequate to describe all data. When expressing it as a function of \( x_T \) and \( \sqrt{s} \), \( n \) was observed to decrease as \( x_T \) or \( \sqrt{s} \) increased, reaching values around 5 at the top of the explored \( p_T \) and \( \sqrt{s} \) range.

Following a suggestion by Bjorken\(^2\), experiments were also performed using calorimeters in order to trigger on the whole jet, so as to avoid the trigger bias mentioned above. The first experiments using this technique were performed at Fermilab\(^1\,\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!
was subdivided into 240 independent cells, each covering $15^\circ$ in $\phi$ and $9^\circ$ in $\theta$. For a given amount $E_i$ of energy deposited in cell $i$, a transverse energy $E_T^{(i)} = E_i \sin \theta_i$ was defined, where $\theta_i$ is the polar angle of the cell centre, and three different triggers were formed: the first one used the sum of the $E_T^{(i)}$ over the whole calorimeter ($\Delta \phi = 2\pi$); the second one extended this sum to the cells contained in two azimuthal sectors of $90^\circ$ each ($\Delta \phi = 2 \times 90^\circ$); and the third one was limited to one such sector only ($\Delta \phi = 90^\circ$). If at high transverse energies the events selected by these triggers were dominated by two coplanar high-$p_T$ jets, the event rates would be roughly in the ratio 2:1:1 for the three trigger types, respectively.

Figure 4 shows the total transverse energy spectra measured for the three different trigger types just mentioned and for different incident energies and particle types. The main feature of these results is the increase of the cross-section at large $E_T$ by about two orders of magnitude for a doubling of the trigger azimuthal coverage, in strong disagreement with two-jet dominance.

In order to search for a two-jet structure, each event was projected onto a plane perpendicular to the beam direction and a search was made for the axis which maximizes the sum of the squares of all transverse momenta (the major axis). If $a$ is the value of such a sum, and $b$ is the value corresponding to the minor axis, a

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig4.png}
\caption{Distribution of the total transverse energy measured in the NA5 calorimeter for three different azimuthal intervals.}
\end{figure}
The parameter called "planarity" can be defined for each event as

\[ P = \frac{a - b}{a + b}, \]

which is zero for events with perfect azimuthal symmetry, and equal to unity for two infinitely collimated jets.

The result of this analysis on the NA5 events shows no increase of the average planarity up to total transverse energies as high as 70% of the allowed maximum. Furthermore, the shape of the planarity distribution disagrees with the shape expected in the case of hard parton scattering, while approximately reproducing the results of a calculation based on a low-\( p_T \) model.

The same conclusions have been reached by a similar experiment performed at Fermilab\(^{25}\).

These results are in sharp contrast with the case of \( e^+e^- \) annihilation into hadrons, where jets are obvious above \( \sqrt{s} = 10 \text{ GeV} \). They have been interpreted either as the effect of multiple gluon bremsstrahlung from the initial-state partons participating in the hard collision\(^{28}\); or as the effect of the tails of the multiplicity distributions in ordinary hadronic collisions at low \( p_T \)\(^{27}\). In either case, they have cast doubts on the usefulness of a total transverse energy trigger as a way to obtain clean samples of high-\( p_T \) jets which are free of the biases introduced by previous experiments. This feeling of discouragement persisted even after the first physics run at the \( pp \) Collider (December 1981) had been completed. In an article which appeared in the February 1982 issue of Physics Today\(^ {28}\), reporting the first preliminary results from the Collider experiments, it can be read: "... the anomalously high total transverse energy appears generally to be distributed quite uniformly among the particles emerging in all azimuthal directions. Clean parton-model jets will be much more elusive in hadron-hadron scattering than in \( e^+e^- \) collisions"\(^ {11}\).

We shall see in Section 4 how this pessimistic view has been contradicted by the dramatic emergence of unambiguous jets in events with total transverse energies in excess of \( \sim 50 \text{ GeV} \) at the \( pp \) Collider.

3. THE CERN \( pp \) COLLIDER FACILITY AND ITS TWO MAJOR EXPERIMENTS

3.1 The machine

The discovery of the weak neutral current interaction in 1973\(^ {29}\) was the first experimental support for the electroweak theory\(^ {30}\). By measuring the weak mixing angle it became possible to predict the mass values of the weak IVBs within the framework of the standard model\(^ {31}\). Since these values were \( \sim 80 \) and \( \sim 90 \text{ GeV}/c^2 \) for the charged and the neutral IVB, respectively, these particles could not be produced at any of the high-energy accelerators existing at that time (the highest collision energy, \( \sqrt{s} = 63 \text{ GeV} \), was reached at the CERN ISR). A new machine was needed, therefore, to verify experimentally the existence of these particles with the properties predicted by the standard model.
The easiest way to achieve the necessary energy was to modify one of the existing proton accelerators (the CERN SPS or the equivalent machine at Fermilab) so that it could be operated as a collider, namely with a proton and an antiproton beam circulating in opposite directions in the same ring and undergoing head-on collisions [pp collisions were considered instead of $e^+e^-$ because for the latter case the existing ring was not large enough to achieve the necessary energy, given the very high energy loss by synchrotron radiation suffered by the electrons].

However, the disadvantage of a pp collider is that it is very difficult to obtain $\bar{p}$ beams with particle densities high enough to achieve the collision rate necessary to produce and detect the IVBs. More precisely, since the cross-section to produce these particles is of the order of a few nanobarns (see Section 5), it is necessary to achieve luminosities of at least $10^{28}$ cm$^{-2}$ s$^{-1}$ in order to produce a few IVBs per day of collider operation (the luminosity $L$ is defined as the ratio between the event rate for a given process and its cross-section). Because of the low $\bar{p}$ yield in $p$-nucleus collisions (of typically $\sim 10^{-6}$ $\bar{p}$'s per incident proton), it is necessary to accumulate in a special storage ring the $\bar{p}$'s produced over several hours. However, although sufficient in number, these $\bar{p}$'s are characterized by a very low phase-space density which results from their diffused production in proton-nucleon collisions. It is necessary, therefore, to apply in addition a "cooling" technique, which increases the $\bar{p}$ phase-space density to values comparable with those of the proton beam.

Two beam-cooling techniques have been developed:

i) **Electron cooling** uses intense well-collimated and monoenergetic electron beams to cool disordered $\bar{p}$ beams whose average velocity is equal to that of the electrons. This technique was invented in 1966 by Budker and applied successfully to proton beams in the early seventies.

ii) **Stochastic cooling** uses the fluctuations of the average beam position in phase space, which occur because the beam contains a finite, though large, number of particles. These fluctuations are measured by means of suitable pick-up electrodes located at various positions along the accumulator ring, and they are used to generate correction signals which are applied to the beam in other points of the ring. Since only the average beam position in phase space is corrected in this way, and not the individual particles, this method must be applied to the circulating beam for a time of the order of several minutes in order to be effective. This technique was invented by van der Meer in 1972 and applied successfully to the protons circulating in the CERN ISR in 1974. We note that stochastic cooling does not violate Liouville's theorem, which states the conservation of phase-space density for particle beams under the effect of external electromagnetic fields, because this theorem is valid only for continuous media with an infinite number of degrees of freedom and, as mentioned above, this technique does not work for beams with an infinite number of particles.
It was suggested in 1976 by Rubbia et al.\textsuperscript{37} that either one of these cooling techniques could be used to obtain a p beam with intensity and particle density high enough to achieve the luminosity necessary to produce the IVBs with reasonable rates in the CERN SPS (or the Fermilab machine) suitably modified for collider operation. Following this suggestion, an Initial Cooling Experiment (ICE) was performed at CERN in a magnetic ring which had been previously used for an experiment to measure the muon $g-2$ factor. ICE demonstrated the validity of both cooling techniques with p and $\bar{p}$ beams: in particular, it achieved successful stochastic cooling of both betatron oscillations and momentum spread\textsuperscript{38,39}. A full pp project based on stochastic cooling was approved at CERN in June 1978, following the encouraging results from ICE\textsuperscript{40}.

The operation of the CERN pp collider consists of four stages: $\bar{p}$ accumulation, beam injection into the SPS, acceleration, and coasting.

During $\bar{p}$ accumulation, a beam of $\sim 10^{13}$ protons is accelerated to 26 GeV in the Proton Synchrotron (PS), extracted from this machine, and transported to the $\bar{p}$ production target. Antiprotons produced with a momentum of 3.5 GeV/c (where the yield is maximum) are focused by a magnetic horn located just after the target, and injected into the Antiproton Accumulator (AA) (see Fig. 5). The AA is a fixed-field machine with an unusually large aperture, which is needed in order to collect the largest possible p phase space. Fast precooling of the injected $\bar{p}$ beam occurs in the outer part of the vacuum chamber, reducing the momentum spread from 1.5% to 0.2% in $\sim 2$ s. The precooled beam is then decelerated and added to the top of the already existing $\bar{p}$ stack, which is continuously cooled both transversally and longitudinally. The injection orbits and the stack region are separated by movable ferrite shutters, which are needed to isolate the cooling from the precooling system. This whole operation, from proton acceleration in the PS to the deposition of the newly injected $\bar{p}$'s in the stack, is repeated once every 2.4 s. Typical stacking rates of $\sim 5 \times 10^9$ $\bar{p}$/h have been achieved, corresponding to $\sim 10^{11}$ $p/d$. Cooling in the AA reduces the product of phase space and momentum bite occupied by the $\bar{p}$'s by a factor of $\sim 10^8$ which is needed in order to match the acceptance of the SPS.

When a sufficiently dense $\bar{p}$ beam has been accumulated in the AA, beam injection into the SPS is achieved using six consecutive PS cycles (the time interval between cycles is 2.4 s). Firstly, three proton bunches, each containing $\sim 10^{11}$ protons, are accelerated in the PS to an energy of 26 GeV and injected into the SPS through transfer line TT10 (see Fig. 5). In the three following cycles, three $\bar{p}$ bunches of typically $10^{10}$ $\bar{p}$ each, belonging to the densest part of the $\bar{p}$ beam in the AA, are extracted from the AA and injected into the PS through transfer line TTL2 (Fig. 5). Here they are accelerated to 26 GeV in a direction opposite to that previously used for the protons, and they are then injected into the SPS through transfer line TT70 (Fig. 5). The relative injection timing of the bunches is controlled with a precision of $\sim 1$ ns to ensure that bunch crossing occurs in the centre of the experimental areas.
After injection the three $p$ and the three $\bar{p}$ bunches, which occupy equidistant positions along the SPS circumference, are simultaneously accelerated by the SPS radio-frequency (RF) system up to an energy of 270 GeV, which represents the present limit imposed on collider operation by power dissipation at constant magnetic field.

After acceleration the beams are kept bunched by the RF system. Each bunch has an approximately Gaussian shape with a longitudinal r.m.s. of $\sim 20$ cm and transverse dimensions of the order of 1 mm. Although the bunches cross in each of the six long straight sections of the SPS ring, LSS1-LSS6, experiments are performed only in LSS4 and LSS5, where two large underground halls have been excavated. Special quadrupoles are provided in these two regions to squeeze locally the beam transverse dimensions and thus increase the luminosity.

The highest luminosity achieved so far in a physics run is $1.5 \times 10^{29} \text{ cm}^{-2} \text{ s}^{-1}$, which corresponds to $\sim 9000$ $p\bar{p}$ collisions per second. The luminosity decays exponentially with time, with an average decay time of $\sim 18$ h. This value also represents the typical length of a physics run at the Collider. At the end of it, the circulating beams are dumped and new beams are injected.
The yoke of the magnetic spectrometer is laminated and scintillator is inserted between the iron plates to form the central hadronic calorimeter which surrounds both e.m. calorimeters. The hadronic calorimeter (see Figs. 6c and 6d) consists of 450 independent cells, with typical size $\Delta \theta \times \Delta \phi$ equal to $15^\circ \times 18^\circ$ in the central region and $5^\circ \times 10^\circ$ in the forward regions, respectively. Their thickness is $\sim 5$ and $\sim 7$ absorption lengths, respectively. The energy resolution is typically $\sigma_E/E \approx 0.8 / \sqrt{E}$ ($E$ in GeV).

Muon detectors surround the magnet yoke (see Fig. 6a). They consist of two planes of drift chambers covering the polar angle interval $5^\circ < \theta < 175^\circ$ and the full azimuth, separated by a distance of 60 cm. Each of these planes is made of 50 drift chambers of $4 \times 6$ m$^2$ size. Each chamber consists of four layers of drift tubes (maximum drift space 7 cm), which define two orthogonal coordinates. By knowing the position of the interaction vertex and the muon-track parameters in these chambers, it is possible to obtain an independent measurement of the muon momentum with a relative precision of $\sim 20\%$ through the determination of its deflection in the magnet yoke.

In addition to the detectors just described, more calorimeters and track detectors located along the beam pipe on both sides of the spectrometer cover the region of polar angles from $5^\circ$ down to $\sim 0.2^\circ$ with respect to the beam line. A schematic view of the overall detector configuration is shown in Fig. 6d.

3.3 The UA2 experiment

The UA2 experiment was designed mainly to search for the weak bosons by identifying their decay modes into electrons, and to study final states containing high-$p_T$ jets$^{11}$. The detector (Fig. 7a) consists of three parts:

- The vertex detector, which is a system of cylindrical chambers to reconstruct charged particle tracks in a region without magnetic field. This detector covers the polar angle interval $20^\circ < \theta < 160^\circ$. It consists of four multiwire proportional chambers (MWPCs) and two drift chambers, all with wires parallel to the beam axis, providing a total of 16 points per track. Reconstruction in three dimensions is achieved in the MWPCs by measuring the charge induced on cathode strip at $\pm 45^\circ$ to the wires; and in the drift chambers using the technique of charge division. The drift chambers provide also a measurement of the specific ionization for each charged particle.

The vertex detector contains also a counter hodoscope made of 24 scintillator strips parallel to the beam axis. A "preshower" counter, consisting of a 1.5 r.l. thick tungsten cylinder followed by a MWPC with cathode strip read-out and pulse-height measurement on the wires, is located just behind the last chamber, covering the central region $40^\circ < \theta < 140^\circ$. As we shall see in Section 4, this device is essential for electron identification.
The central calorimeter, which is a system of 240 independent counters covering the full azimuth (see Fig. 7b) and the polar angle interval $40^\circ < \theta < 140^\circ$. Each counter has an angular acceptance $\Delta \theta \times \Delta \phi = 10^\circ \times 15^\circ$ and it consists of a first e.m. section (lead-scintillator) 17 r.l. thick, followed by two independent hadronic sections (iron-scintillator), each two absorption lengths thick. Light is collected by wavelength shifting plates located on two opposite sides of each counter. The energy resolution is $\sigma/E \approx 0.19/\sqrt{E}$ for e.m. showers, and $\approx 0.32 E^{-1/4}$ for hadronic showers.

Fig. 7 Views of the UA2 detector: a) cross-section in the vertical plane containing the beam axis; b) cross-section of the central detector normal to the beam axis; c) exploded view of a sector in one of the forward detectors.
(E in GeV). All counters have been calibrated using 10 GeV electron and muon beams from the CERN PS; frequent checks with an intense $^{60}$Co source and other devices are made to monitor the calibration stability.

Until the end of 1982, the azimuthal range of this calorimeter was 300° only. The remaining interval ($\pm 30°$ around the horizontal plane) was covered by a magnetic spectrometer used to perform measurements which are not relevant to the subject of these lectures.

The two forward detectors, covering the polar angle intervals 20° < $\theta$ < 37.5° and 142.5° < $\theta$ < 160° and the full azimuth. Each detector consists of twelve sectors in which a toroidal magnetic field is generated by twelve coils equally spaced in azimuth (the field integral is $\approx 0.38$ T•m). Following the magnetic field volume, each sector contains nine drift chamber planes, which are used to measure the charged particle momenta together with the information from the vertex detector. A preshower counter, consisting of a 1.5 r.l. thick lead-iron plate followed by four layers of proportional tubes, is located after these chambers. Energy measurement, limited to e.m. showers only, is performed with a calorimeter containing ten independent counters per sector ($\Delta \theta \times \Delta \phi = 2.5° \times 15°$ per counter). Each counter is a lead-scintillator multilayer sandwich, subdivided in depth into two independent sections, 24 and 6 r.l. thick. Energy resolution is similar to that of the central calorimeter. Calibration and its monitoring are also achieved in a similar way.

A view of a sector of a forward detector is shown in Fig. 7c.

No muon detector is implemented in the UA2 experiment.

4. HIGH-pT HADRONIC FINAL STATES AT THE pp COLLIDER

4.1 Two-jet dominance

The first evidence for clear two-jet structure in events with large transverse energy was presented in 1982 by the UA2 group, soon after the first physics run at the Collider. Events with large transverse energy were selected by a trigger which summed all transverse energies deposited in the UA2 central calorimeter by the outgoing particles, and required this sum $\sum E_T$ to exceed a given threshold. Figure 8 shows the observed $\sum E_T$ distribution from more recent data. For $\sum E_T > 60$ GeV there is a clear departure from exponential, an effect not seen in lower-energy experiments (see Fig. 4).

Energy clusters, as expected in the case of jet production, are constructed by joining all calorimeter cells which share a common side and contain at least 400 MeV. In each event, these clusters are then ranked in order of decreasing transverse energies ($E^1_T > E^2_T > E^3_T > \ldots$). They consist typically of 3 cells for $E_T = 2$ GeV and 10 cells for $E_T = 40$ GeV.

96
The mean values of the fractions $h_1 = E_T^1 / \Sigma E_T$ and $h_2 = (E_T^1 + E_T^2) / \Sigma E_T$ are shown in Fig. 9 as a function of $\Sigma E_T$. Their behaviour illustrates the emergence of a dominant two-jet structure at large values of $\Sigma E_T$. An event containing only two jets of equal transverse energies would have $h_1 = 0.5$ and $h_2 = 1$. Figure 9 shows that, when $\Sigma E_T$ is large enough, a large fraction of $\Sigma E_T$ is shared on the average by two clusters only, with roughly equal transverse energies.

The azimuthal separation $\Delta \phi_{12}$ between the two largest clusters is shown in Fig. 10 for events with $\Sigma E_T > 60$ GeV and $E_T^1, E_T^2 > 20$ GeV. A clear peak at $\Delta \phi_{12} = 180^\circ$ is observed, indicating that the two clusters are back to back in a plane perpendicular to the beam axis, as expected for two-jet production.
The emergence of jet structures in events with large $\Sigma E_T$ is even more dramatically illustrated by inspecting the transverse-energy distribution in the $\Theta-\phi$ plane. Figures 11a, b, c, d show such a distribution for four typical events with $\Sigma E_T > 100$ GeV. The transverse energy is seen to be concentrated within two (or, more rarely, three) small angular regions. Furthermore, reconstructing the charged-particle tracks in these events shows several collimated tracks pointing to the energy clusters (Fig. 12).

Fig. 10 Azimuthal separation between the two clusters having the highest $E_T$ ($E_T > 20$ GeV) for events with $\Sigma E_T > 60$ GeV.

Fig. 11a to d) Four typical transverse-energy distributions for events with $\Sigma E_T > 100$ GeV in the $\Theta-\phi$ plane. Each $\Theta-\phi$ bin represents a cell of the UA2 calorimeter.
Fig. 12 Transverse view of a two-jet event in the UA2 detector. Several reconstructed charged-particle tracks are seen to point to the two energy clusters.

The apparent disagreement between these results and those of NA5, where no evidence for jets was found\(^1\), can easily be explained assuming that the \(\Delta E_T\) distribution consists of two components, the first one being due to hard parton scattering, and the second to the high multiplicity tail of soft, low-\(p_T\) collisions, and estimating their contributions as a function of the collision energy \(\sqrt{s}\).

Fig. 13 Contributions to the total transverse-energy spectrum from the soft component (full line) and the hard component (broken line). The sum of the two contributions is shown as the dash-dotted line. a) \(pp\) at \(\sqrt{s} = 25\) GeV; b) \(pp\) at \(\sqrt{s} = 60\) GeV; c) \(\bar{p}p\) at \(\sqrt{s} = 540\) GeV.
Figure 13 shows the results obtained by Åkesson and Bengtsson for the case of a calorimeter covering the full azimuth and the pseudorapidity interval $\Delta n = \pm 1$ (a very good approximation to the UA2 and NA5 calorimeters). Whereas at $\sqrt{s} = 25$ GeV (the value relevant to the NA5 experiment) the hard component is everywhere smaller than the soft one by at least one order of magnitude (Fig. 13a), at the $p\bar{p}$ Collider the contrary is true for $E_{T} > 30$ GeV (Fig. 13c). At ISR energies (Fig. 13b) hard parton scattering dominates over the soft component only for total transverse energies in excess of $\sqrt{s}/2$, and jets have indeed been observed in these rather difficult conditions.

4.2 Theoretical interpretation

Jet production at the Collider is interpreted in the framework of the parton model as hard scattering among the constituents of the incident proton and antiproton. Since the initial state contains quarks $q$, antiquarks $\bar{q}$, and gluons $g$, there are several elementary sub-processes which can contribute to jet production. For each subprocess the scattering cross-section is given to first order in $\alpha_s$ by the expression

$$\frac{d\sigma}{dt} = \frac{\alpha_s^2}{\hat{s}} |M|^2,$$

where $\hat{s}$ is the square of the total energy in the centre of mass of the two partons, and $t$ is the square of the four-momentum transfer, defined as

$$t = -\frac{\hat{s}}{2} (1 - \cos \theta^*)$$

$\theta^*$ being the scattering angle.

Explicit expressions for $|M|^2$ in Eq. (1) are given in Table 1 for the various sub-processes, as a function of the Mandelstam variables, $\hat{s}$, $t$, and $u$ ($u = -\hat{s} - t$ under the assumption of massless partons). In order to illustrate the relative importance of the sub-processes, Table 1 displays also the numerical values of $|M|^2$ at $\theta^* = 90^\circ$, where $t = u = -\hat{s}/2$. It is evident that terms involving initial gluons, such as $gg$ and $qg$ (or $g\bar{q}$) scattering, become dominant at small values of $x$, where the gluon density in the incident hadrons is comparable to that of the quarks, or larger.

4.3 Inclusive jet yield

It is now possible to compute the leading term of the inclusive jet yield as a sum of convolution integrals:

$$\frac{d\sigma}{dp_T^2 d\eta} = 2\pi p_T^2 \sum_{A,B} \int dx_1 \, dx_2 \, F_A(x_1, Q^2) F_B(x_2, Q^2) \times$$

$$\times \delta(\hat{s} + t + u) \alpha_s^2(Q^2) \sum_{f} \frac{|M|_{AB,f}^2}{\hat{s}},$$

where $M_{AB,f}$ is the matrix element for the subprocess $f$. This expression is the leading term in the inclusive jet yield.
Matrix elements for parton scattering

| Subprocess | $|M|^2$ | $|M|^2$ at $\theta^\circ = 90^\circ$ |
|------------|--------|-------------------------------|
| $qq' \to qq'$ | $\frac{4}{9} \frac{\delta^2 + u^2}{t^2}$ | 2.22 |
| $qq' \to \bar{q}q'$ | $\frac{4}{9} \frac{\delta^2 + \bar{u}^2}{t^2}$ | 0.22 |
| $qq \to qq$ | $\frac{4}{9} \left( \frac{\delta^2 + u^2}{t^2} + \bar{u}^2 \frac{\delta^2 + \bar{t}^2}{u^2} \right) - \frac{8}{27} \frac{u^2}{\delta t}$ | 3.26 |
| $\bar{q} \bar{q} \to q'q'$ | $\frac{4}{9} \frac{t^2 + u^2}{\delta^2}$ | 2.59 |
| $qq \to gg$ | $\frac{32}{27} \frac{u^2 + t^2}{ut} - \frac{8}{3} \frac{u^2 + t^2}{\delta^2}$ | 1.04 |
| $gg \to \bar{q}q$ | $\frac{1}{6} \frac{u^2 + t^2}{ut} - \frac{3}{8} \frac{u^2}{\delta^2}$ | 0.15 |
| $gg \to q\bar{q}$ | $\frac{4}{9} \frac{u^2 + \bar{u}^2}{\delta^2} + \frac{u^2 + \bar{u}^2}{\delta^2}$ | 6.11 |
| $gg \to gg$ | $\frac{8}{27} \left( \frac{\delta^2}{t^2} - \frac{u^2 + \bar{u}^2}{\delta^2} - \frac{\delta t}{u^2} \right)$ | 30.4 |

a) $q$ and $q'$ denote quarks with different flavours.

where $F_A$ and $F_B$ are structure functions describing the parton density in the incident hadrons and the sum extends over all initial parton types $A$, $B$, and all possible final states $f$.

The structure functions $F(x,Q^2)$ are determined in deep inelastic lepton-nucleon scattering experiments ($Q^2 \leq 20 \text{ GeV}^2$) and extrapolated to the $Q^2$ range of the Collider ($\gtrsim 500 - 10^4 \text{ GeV}^2$), according to the predicted QCD evolution.

Figure 14 shows the inclusive jet production cross-section around $\eta = -\ln \tan \theta/2 = 0$ as a function of the jet transverse momentum $P_T$, as measured in the UA1 and UA2 experiments. The additional systematic errors, not shown by the error bars of Fig. 14, are $\pm 60\%$ for the UA1 data, and $\pm 40\%$ for the UA2 data. Also shown in
Fig. 14 Inclusive jet-production cross-section at $\sqrt{s} = 540$ GeV as a function of jet transverse momentum. The shaded band represents QCD predictions.

Fig. 15 The original prediction by Horgan and Jacob$^{53}$ on the increase of the inclusive jet yield between ISR and collider energies.

Fig. 14 is a band of QCD predictions$^{53,56,57}$, whose width serves to illustrate the uncertainties in the theory arising mostly from different parametrizations of the structure functions.

In spite of the experimental and theoretical uncertainties the agreement between data and theory is remarkable, especially because the theoretical curves of Fig. 14 are not a fit to the data but represent an absolute prediction. It is also worth while noting that the jet yield at the Collider is much larger (by more than three orders of magnitude for $p_T > 20$ GeV/c) than that measured at ISR energies$^{50}$. This fact was first pointed out by Horgan and Jacob$^{53}$ and it is illustrated in Fig. 15.
Figure 16 shows separately the contributions to the inclusive jet yield from $gg$, $qg$ (or $\bar{q}g$), and $qq$ final states, as calculated by Antoniou et al.\textsuperscript{56}). It is clear that gluon jets dominate over the $p_T$ range explored in the Collider experiments, in contrast with $e^+e^-$ collisions where the production of quark jets is the main feature of hadronic final states.

4.4 Transverse momentum of the two-jet system

If jet production is the result of hard parton scattering, the two-jet system is expected to have a limited total transverse momentum $p_{T,\text{jj}}$ because the two initial partons have only a small transverse motion within the incident hadrons. The longitudinal momentum $p_{L,\text{jj}}$, on the contrary, may take large values because the colliding partons have, in general, very different longitudinal momenta.

Experimentally $p_{T,\text{jj}}$ is determined from the sum of two large and approximately opposite vectors, and it is sensitive, therefore, to instrumental effects such as the calorimeter energy resolution and incomplete containment due to edge effects. However, these effects can be made small by only considering the component of $p_{T,\text{jj}}$ along the bisector of the angle between the two jet transverse momenta, $p_{T,\text{jj}}^\parallel$ (see Fig. 17a).

Figure 17b shows the $p_{T,\text{jj}}^\parallel$ distribution as measured in the UA2 experiment\textsuperscript{58}) for a sample of two-jet events with both jets well contained within the calorimeter acceptance and each having $p_T > 20$ GeV/c. Also shown in Fig. 17b is a Monte Carlo prediction which takes into account the detector response and assumes that the distribution $dn/d(p_{T,\text{jj}})^2$ is Gaussian in $p_{T,\text{jj}}$. The data are found to be consistent with this hypothesis, and the mean value of $p_{T,\text{jj}}$ is found to be $\langle p_{T,\text{jj}} \rangle = 5 \pm 2$ GeV/c, with no significant dependence on the invariant mass of the two-jet system.
Fig. 17  a) Definition of the component $p_{T}^{jj}$ of the total transverse momentum of the two-jet system; b) $p_{T}^{jj}$ distribution for two-jet events with both jets having $p_{T} > 20 \text{GeV/c}$. The curve is a Monte Carlo prediction (see text).

4.5 Angular distribution of parton-parton scattering

The study of the jet angular distribution offers the possibility of measuring directly the angular distribution of parton-parton scattering. This distribution can be obtained for each individual subprocess, replacing $t$ and $u$ in Table 1 by the corresponding definitions [see Eq. (3)] which contain an explicit dependence on $\cos \theta^{*}$. For example, in the case of the subprocess $gg \rightarrow gg$ one finds

$$
\frac{d\sigma}{d\cos \theta^{*}} \propto 3 - \frac{\sin^{2} \theta^{*}}{4} + \frac{1 + \cos^{2} \theta^{*}}{4 \sin^{2} \theta^{*}} \frac{\cos^{2} \theta^{*}}{2},
$$

(5)

The familiar singularity at $\theta^{*} = 0$ due to the term $\sin^{-4} (\theta^{*}/2)$ is typical of gauge vector boson exchange. In the subprocesses $qg \rightarrow qg$ (or $\bar{q}g \rightarrow \bar{q}g$) and $gg \rightarrow gg$ it arises from the three-gluon vertex which exists because QCD is a non-Abelian theory (see the relevant diagrams in Fig. 18). This singularity is also present in the subprocess $q\bar{q} \rightarrow q\bar{q}$, but in this case it would be present in an Abelian theory as well, as for $e^+e^-$ elastic scattering in QED.

The determination of the cos $\theta^{*}$ distribution from the experimental data is complicated by the fact that the initial parton-parton system is not at rest in the laboratory. Under the assumption that the two jets are massless, the velocity of the centre-of-mass frame of the two outgoing jets is given by

$$
\frac{\mathbf{v}}{\gamma} = \frac{\mathbf{p}_{1} + \mathbf{p}_{2}}{E_{1} + E_{2}},
$$

(6)
where $E_1$ and $E_2$ are the cluster energies, as measured in the calorimeter, and $p_1$ and $p_2$ are the jet momenta, with magnitudes equal to $E_1$ and $E_2$, respectively, and orientations along the lines joining the event vertex with the cluster centroids.

A Lorentz transformation along $\vec{\ell}$ transforms the two jet momenta into a back-to-back configuration. In the case $p_T^{jj} = 0$ their angle with respect to the beam axis is equal to the parton-parton scattering angle $\theta^*$. When $p_T^{jj} \neq 0$ it is not possible to define $\theta^*$ unambiguously because the two incident partons may have different transverse motions. In this case we choose to approximate $\theta^*$ by the angle between the jet momenta and the external bisector of the angle defined by the incident $p$ and $\bar{p}$ momenta in the centre-of-mass frame of the two jets (this is the so-called Collins-Soper convention), which corresponds to the assumption that $p_T^{jj}$ is equally shared between the two incident partons.

Figure 19 shows the $\cos \theta^*$ distribution measured in the UA1 experiment for jets with $p_T > 20$ GeV/c produced in the pseudorapidity range $|\eta| < 2.5$. Both the experimental data and the theoretical curves for the three dominant subprocesses, also shown in Fig. 19, are normalized to 1 at $\cos \theta^* = 0$. The data are in good agreement with QCD expectations and they clearly show the increase towards the forward directions expected from the $\sin^{-6} (\theta^*/2)$ singularity.

Similar conclusions have been reached in the UA2 experiment, where however the $\cos \theta^*$ distribution is only measured in the range $|\cos \theta^*| < 0.6$ because of the smaller polar angle interval covered by the UA2 central calorimeter.
4.6 Jet fragmentation

A high-$p_T$ jet consists in general of many particles (fragments) whose motion with respect to the jet axis (defined as the direction of the jet total momentum $P$) can be described using two variables: the fractional longitudinal momentum $z = p \cdot \hat{P}/P^2$, where $p$ is the momentum of the fragment being considered (obviously $0 < z < 1$); and the component of $p$ perpendicular to the jet axis, $q_T^2$. The $z$ distribution, $D(z)$, is called the jet fragmentation function.

In order to determine experimentally the function $D(z)$ for a particular type of jet fragments, it is necessary to measure the momentum of each individual fragment. This is only possible in the UA1 detector, where charged-particle momenta can be measured in the large magnetic field volume.

A complication in the determination of $D(z)$ arises from the presence in the final state of particles which do not belong to the jet but are associated with the initial partons which did not take part in the hard collision. These particles are generally referred to as "spectators" and the system of all spectators in a hard collision is called the underlying event. Since the spectators have, in general, low $p_T$ values, they populate the low-$z$ region and the function $D(z)$ can be reliably determined, therefore, only above a certain value of $z$ which decreases with increasing $P$. 

Fig. 19 Distribution of $\cos \theta^*$ for hard parton scattering, deduced from two-jet events with $p_T > 20 \text{ GeV/c}, |\eta| < 2.5$. 

---

106
This complication is absent, of course, for jets produced in $e^+e^-$ annihilation, where the final state contains no spectators.

In the UA1 experiment\(^{60}\) only jets with $P > 30$ GeV/c are considered and all the charged particles with relatively high $z$ values are found to be contained in a $35^\circ$ half-aperture cone around the jet axis. However, particles with $z < 0.02$ have a rather isotropic distribution with respect to the jet axis and their association with the jet is not obvious.

Figure 20 shows $D(z)$ for charged fragments, as measured in UA1 for $z > 0.02$. This distribution is seen to fall rapidly with increasing $z$, approaching an exponential behaviour at large $z$ values.

It is interesting to compare the form of $D(z)$ found at the pp Collider with that measured for jets produced in $e^+e^-$ annihilation. This comparison is also made in Fig. 20, which displays data obtained by the TASSO Collaboration\(^{61}\) in $e^+e^-$ collisions at $\sqrt{s} = 34$ GeV, corresponding to jet momenta of 17 GeV/c. Although the jet momenta are different for the two cases, the comparison is meaningful because scaling-violation effects on $D(z)$ are known to be small\(^{62}\). Figure 20 shows no difference within errors between the two distributions. Here one must remember that jets from $e^+e^-$ annihilation arise from quark fragmentation (the elementary subprocess in this case is $e^+e^- \rightarrow q\bar{q}$), whereas jets at the Collider are dominated by gluon fragmentation, as discussed in Section 4.3. The data of Fig. 20 show that the two fragmentation functions are very similar, at least for $z > 0.02$. 

![Fig. 20 Charged-particle fragmentation function for jets with $E_T > 30$ GeV at the pp Collider, compared with similar results from the TASSO detector at PETRA ($e^+e^-$ collisions, $\sqrt{s} = 34$ GeV).](image)
The UA1 results on the transverse motion of the fragments with respect to the jet axis are shown in Fig. 21, where the invariant q_T spectrum is shown together with the result of a fit to the form

$$\frac{1}{q_T} \frac{dN}{dq_T} = \frac{A}{(q_T + q_0)^N},$$

which was found to describe well the charged-particle transverse momentum distribution in minimum-bias events. The best-fit parameters in Eq. (7) have values $q_0 = 4$ GeV/c, $N = 14.8$.

Only jets with $p_T > 30$ GeV/c are considered in the data of Fig. 21. Fragments with $z > 0.1$ have a mean $q_T$ value of 600 MeV/c, increasing to 700 MeV/c if only jets with $p_T > 50$ GeV/c are considered.

### 4.7 Charged-particle multiplicity in jets

The average charged-particle multiplicity in a jet, $\langle n_{ch} \rangle$, can be obtained by knowing the fragmentation function $D(z)$ at all $z$ values:

$$\langle n_{ch} \rangle = \int_0^1 D(z) \, dz.$$  

This procedure cannot be applied to the data of Fig. 20 because $D(z)$ is not measured for $z < 0.02$ and the uncertainty in the extrapolation to $z = 0$ is too large to make the result meaningful.
A method to overcome this difficulty has been attempted by the UA2 Collaboration\(^{48}\). Events are considered in which two jets, each with \(p_T > 15\) GeV/c and separated in azimuth by at least 150\(^\circ\), are observed in the central calorimeter (40\(^\circ\) < \(\theta\) < 140\(^\circ\)). Figure 22 shows the track density \(\rho\) as a function of the azimuthal separation \(\Delta\phi\) between each charged-particle track reconstructed in the plane perpendicular to the beam over the interval 20\(^\circ\) < \(\theta\) < 160\(^\circ\) and the axis of the higher \(p_T\) jet. Two distributions are shown, corresponding to two different bands of \(M_{jj}\), the invariant mass of the two-jet system calculated assuming massless jets. The distribution relative to the higher \(M_{jj}\) values is more peaked, showing that higher-energy jets are more collimated.

The average charged-particle multiplicity in a jet is obtained from the distributions of Fig. 22 using the relation

\[
\langle n_{\text{ch}} \rangle = \frac{1}{2} \int_0^{\pi} \left[ \rho(\Delta\phi) - \rho_0 \right] d(\Delta\phi),
\]

where \(\rho_0\) is the density of charged spectators. The dependence of \(\langle n_{\text{ch}} \rangle\) on \(M_{jj}\) is shown in Fig. 23 under the assumption that \(\rho_0\) is constant over the entire \(\Delta\phi\) range and its value is given by \(\rho(\pi/2)\). These data represent a lower bound to the real multiplicities because we have assumed that the density of jet fragments with \(\Delta\phi = \pi/2\) is zero.

Figure 23 shows also the mean charged-particle multiplicities for jets produced in \(e^+e^-\) annihilation\(^{64}\), modified according to Eq. (9) to be consistent with the \(p\bar{p}\) data. In \(e^+e^-\) collisions the variable \(\sqrt{s}\) is equivalent to \(M_{jj}\) for the \(p\bar{p}\) case.

---

Fig. 22 Track density as a function of the azimuthal separation \(\Delta\phi\) between each charged-particle track and the axis of the higher-\(p_T\) jet in two-jet events. Two different ranges of \(M_{jj}\), the invariant mass of the two-jet system, are shown.
Calculations based on QCD predict that gluon jets have a higher multiplicity than quark jets\(^65\), and, as already mentioned, jets at the \(p\bar{p}\) Collider are mainly gluon jets whereas \(e^+e^-\) annihilations produce quark jets. This prediction cannot be tested by the data of Fig. 23 because there is no overlap between the intervals of \(\sqrt{s}\) and \(M_{jj}\) for which \(e^+e^-\) and \(p\bar{p}\) data are available.

4.8 Multijet events

It is known that hadronic final states from \(e^+e^-\) annihilation at high energy exhibit sometimes a clear three-jet structure, which is interpreted as an effect of gluon bremsstrahlung radiated by one of the two outgoing quarks\(^66\)). Such an effect is also expected in the case of \(p\bar{p}\) collisions, where both quarks and gluons produced with high \(p_T\) may radiate other gluons. In particular, the process \(g + g\), which occurs because of the presence of the three-gluon vertex in QCD, has a rate \(9/4\) times higher than that of \(q + q^*\).\(^67\)

An important difference between \(e^+e^-\) and \(p\bar{p}\) collisions is that in the latter case gluons can also be radiated by the initial-state partons as well as at the parton scattering vertex directly. We expect such radiation to be much less correlated with the two main jets than in the case of gluons radiated directly by the two outgoing high-\(p_T\) partons.

It has been observed in the UA2 experiment\(^68\) that the total transverse energy measured in the central calorimeter in addition to the two main jets, \(\sum E_T\), is larger, on the average, than that measured in minimum-bias events. For a total azimuthal acceptance of \(300^\circ\) (this was the configuration of the UA2 central calorimeter in
In 1982, see Section 4.3), the mean value of $\langle \vec{E}_T \rangle$ for events containing two jets with $p_{T1} + p_{T2} > 60 \text{ GeV/c}$ is 13.1 GeV, whereas it is only 5.5 GeV in minimum-bias events.

It has also been observed in UA2 that the underlying event has a total transverse momentum which partially balances the total transverse momentum $p_{T}^{\text{jj}}$ of the two-jet system, as expected.

Approximately 30% of the events with two jets having $p_{T1} + p_{T2} > 60 \text{ GeV/c}$ are found in UA2 to contain an additional energy cluster with a transverse energy in excess of 4 GeV, whereas only 4.5% of minimum-bias events contain such a hard cluster. If we label the two main jets 1 and 2 according to the relation $p_{T1} > p_{T2}$ we can study the correlation between the third cluster and the two main jets by looking at the angle $\omega_{13}$ ($\omega_{23}$) defined by the momentum vector of jet 1 (jet 2) and that of jet 3.

The distributions of $\cos \omega_{13}$ and $\cos \omega_{23}$ are shown in Fig. 24. They exhibit a clear correlation between jet 2 and the third cluster, as expected if one assumes that the third cluster is due to a gluon jet radiated by parton 2. The drops near $\cos \omega = \pm 1$ are instrumental effects resulting from the inability of the detector to resolve jets if they are separated by less than $\sim 30^\circ$.

If the transverse energy of the third cluster is increased, the angular correlation with jet 2 becomes stronger but the event sample is reduced.

All of these observations are in qualitative agreement with the expectations based on gluon bremsstrahlung.

Fig. 24 Angular correlation in three-jet events with $p_{T1} + p_{T2} > 60 \text{ GeV/c}$, $p_{T3} > 4 \text{ GeV/c}$.
As in the case of the $W^\pm$, the $Z^0$ decays mainly into a fermion-antifermion pair $ff$ ($f = \nu$, $e^-$, $\mu^-$, $\tau^-$, or any quark $q$). The coupling of the $Z^0$ to such a pair may be parametrized by an effective Lagrangian given by

$$\mathcal{L}_{\text{eff}} = - \frac{G_F}{\sqrt{2}} \bar{f} Y_f (V_f - A_f Y_f) f Z^0,$$

where

$$V_f = A_f = 1 \text{ for neutrinos;}$$

$$A_f = \pm 1, \quad Y_f = \pm (1 - 4|Q_f| \sin^2 \theta_W) \text{ for fermions } f \text{ with charge } Q_f \geq 0.$$

We note that for charged leptons ($|Q_f| = 1$) the vector coupling $V_f$ is very small because $\sin^2 \theta_W$ has a value close to 0.25.

The partial decay widths into a lepton-antilepton pair $\bar{f}f$ are then given by

$$\Gamma(\bar{f}f) = \frac{G_F M_Z^2}{24\pi \sqrt{2}} (v_f^2 + A_f^2),$$

and those relative to $Z^0$ decay into a $q\bar{q}$ pair:

$$\Gamma(q\bar{q}) = \frac{3G_F M_Z^2}{24\pi \sqrt{2}} (v_q^2 + A_q^2),$$

where, as usual, the factor of 3 results from the three possible colour states of the quarks. Quark-antiquark pairs are expected to belong to the same generation because flavour changing is suppressed in the weak interaction between neutral currents.

Quantitatively one finds

$$\Gamma(\bar{\nu}\nu) : \Gamma(e^+e^-) : \Gamma(\bar{\nu}\nu) : \Gamma(\bar{\nu}\nu) =$$

$$= 2 : 1 + (1 - 4 \sin^2 \theta_W)^2 : 3\left[1 + \left(1 - \frac{8}{3} \sin^2 \theta_W\right)^2\right] : 3\left[1 + \left(1 - \frac{4}{3} \sin^2 \theta_W\right)^2\right],$$

and similar relations for the other two fermion generations. It must be noted, however, that in the decay $Z^0 \to \bar{e}e$ the effect of the large top-quark mass $M_t$ introduces a suppression factor which cannot be neglected.

Setting $\sin^2 \theta_W = 0.215$ one obtains

$$\Gamma(\bar{\nu}\nu) : \Gamma(e^+e^-) : \Gamma(\bar{\nu}\nu) : \Gamma(\bar{d}d) = 2 : 1.02 : 3.55 : 4.53,$$

from which, neglecting the effect of the top-quark mass, we can calculate the branch-
ing ratio for the decay mode $Z^0 \rightarrow e^+e^-$ under the assumption of three fermion generations ($N_G = 3$):

$$\frac{B_{ee}}{\Gamma_Z} = \frac{\Gamma(e^+e^-)}{\Gamma_Z} = \frac{\Gamma(\mu^+\mu^-)}{\Gamma_Z} \approx \frac{1.02}{11.1 N_G} \approx 3\%,$$  \hspace{1cm} (24)

where $\Gamma_Z$ is the total width of the $Z^0$. Using the numerical value $\Gamma(e^+e^-) = 90$ MeV, which can easily be obtained from Eq. (20), the value of $\Gamma_Z$ turns out to be of the order of 3 GeV/c$^2$.

The measurement of $\Gamma_Z$ can be considered as a method to estimate the total number of neutrino types with $m_\nu \ll M_Z/2$, because each decay channel $Z^0 \rightarrow \nu\bar{\nu}$ contributes $\approx 180$ MeV to $\Gamma_Z$ [see Eq. (20)]. Before the collider experiments an upper bound to such a number from particle physics experiments was $N_\nu \lesssim 10^5$. As we shall see later in Section 5.10, this number has been dramatically reduced by the first observation of the $Z^0$ decaying into $e^+e^-$ at the pp collider.

5.4 Production of the IVBs

The production of $W^\pm$ and $Z^0$ in hadron collisions is expected to occur as the result of $q\bar{q}$ annihilation, also known as the Drell-Yan mechanism. In the case of $pp$ collisions, and considering only the valence quarks, the basic subprocesses are

$$\begin{align*}
    u + \bar{d} &\rightarrow W^+ \\
    d + \bar{u} &\rightarrow W^- \\
    u + \bar{u} &\rightarrow Z^0 \\
    d + \bar{d} &\rightarrow Z^0.
\end{align*} \hspace{1cm} (25)$$

Since the transverse momenta of the initial quarks are known to be small, the IVBs are produced with low $p_T$ and the initial $x$-values, $x_1$ and $x_2$, obey the relation $x_1 x_2 s = M^2$, where $M$ is the IVB mass.

The subprocess $u + \bar{d} \rightarrow W^+$ has a cross-section whose value is given by

$$\hat{\sigma}(x_1, x_2) = \sqrt{2\pi G_F M_W} \cos^2 \theta_c \delta(x_1 x_2 s - M^2),$$  \hspace{1cm} (26)

where we have neglected the $W^+$ width. To obtain the inclusive cross-section for $W^+$ production we must perform a convolution integral, which takes into account the $x$ distribution (or structure functions) of the $u$ and $d$ quarks in the incident hadrons, and sum over all relevant subprocesses.

In the case of $pp$ collisions, and considering only valence quarks, we obtain

$$\sigma(pp \rightarrow W^+ + ...) = \int \frac{dx_1}{x_1} \int_0^{1-x_1} dx_2 \hat{\sigma}(x_1, x_2) W^+(x_1, x_2)$$

$$= \sqrt{2\pi G_F} \cos^2 \theta_c \int_0^1 \frac{dx}{x} W^+(x, \frac{1}{x}),$$  \hspace{1cm} (27)

115
$\tau = \frac{M^2}{s}$ and

$$W^+(x_1, x_2) = \frac{1}{3} \left[ U(x_1) D(x_2) + D(x_1) U(x_2) \right] .$$  \hspace{1cm} (28)

$U(x)$ and $D(x)$ being the $x$ distributions of the $u$ and $d$ quarks in the incident protons, which are identical, of course, to those of the $\bar{u}$ and $\bar{d}$ quarks in the incident anti-protons.

In complete analogy, the cross-section for inclusive $Z^0$ production in $pp$ collisions is given by

$$\sigma(pp \rightarrow Z^0 + \ldots) = \frac{1}{\tau} \int \frac{dx}{x} \frac{1}{2} \left( 1 - \frac{2}{3} \sin^2 \theta_W^+ \right) \sin^2 \theta_W^- .$$  \hspace{1cm} (29)

Estimates of the cross-sections for inclusive production of $W^+$ and $Z^0$, based on Eqs. (27) and (29), have been made since 1977 using structure functions with no scaling violation (78).

Figure 25 shows the more recent estimate by Paige (79) for both $p\bar{p}$ and $pp$ collisions, as a function of the collision energy $\sqrt{s}$. Here scaling violations are taken into account, as well as the contributions of $q$ and $\bar{q}$ from the sea. We note that $W^+$ and $W^-$ are produced with equal cross-sections in $pp$ collisions, whereas this is not true in the $p\bar{p}$ case.

The corresponding estimates for inclusive $Z^0$ production are shown in Fig. 26. At the $p\bar{p}$ Collider ($\sqrt{s} = 540$ GeV) the numerical values are

$$\sigma(p\bar{p} \rightarrow W^+ + \ldots) = \sigma(p\bar{p} \rightarrow W^- + \ldots) \approx 1.8 \times 10^{-33} \text{ cm}^2$$  \hspace{1cm} (31)

$$\sigma(p\bar{p} \rightarrow Z^0 + \ldots) \approx 10^{-33} \text{ cm}^2 .$$  \hspace{1cm} (32)

Non-leading QCD terms, represented by the graphs of Fig. 27, have the effect of multiplying these cross-section values by a factor $k$. Such an effect has been observed in the production of lepton pairs by hadron collisions at lower energies, where the value $k = 2$ has been measured (80), and the dilepton transverse momentum has been found to increase with $\sqrt{s}$ (81).

For inclusive $W^\pm$ and $Z^0$ production at the $p\bar{p}$ Collider the value $k = 1.5$ has been predicted (82). Another effect of the non-leading QCD terms is that sometimes...
Fig. 25  Cross-section for inclusive \( W^\pm \) production as a function of \( \sqrt{s} \) for pp and \( \bar{p}p \) collisions (from Ref. 79).

Fig. 26  Cross-section for inclusive \( Z^0 \) production as a function of \( \sqrt{s} \) for pp and \( \bar{p}p \) collisions (from Ref. 79).

Fig. 27  Diagrams describing non-leading QCD corrections to the production of IVBs in hadron-hadron collisions.

the IVBs are produced with high transverse momentum, which is balanced by a hadronic jet (see graphs b to e in Fig. 27).
5.5 Detection of the IVBs

We have seen in the previous section that the IVBs are produced, in general, with a rather low transverse momentum. Their decay products, however, if emitted not very close to the beam direction, will have high $p_T$ because of the large amount of energy released in the decay.

For both $W^+$ and $Z^0$, a large fraction of the decay modes (about 75%) consist of $q\bar{q}$ pairs which appear as two high-$p_T$ hadronic jets. In searching for such configurations we expect a background from the continuum of high-$p_T$ two-jet events discussed in Section 4. As a measure of this background we integrate the inclusive jet cross-section shown in Fig. 14 over $\pm 1$ unit of rapidity and for jet transverse momenta $p_T > 30$ GeV/c. The result is $\sigma_{\text{jet}} \approx 2 \times 10^{-31}$ cm$^2$, to be compared with the expected cross-sections for $W^\pm$ and $Z^0$ production, which are of the order of $10^{-33}$ cm$^2$ [see Eqs. (31) and (32)].

Such an unfavourable signal-to-noise ratio cannot be easily reduced experimentally because there is no way to distinguish jets from $W^\pm$ and $Z^0$ decay from those which result from hard parton scattering. A measurement of the two-jet invariant mass with excellent resolution (of the order of the natural width of the IVBs) would certainly help to identify a peak superimposed on the smooth mass distribution of the background. However, in both UA1 and UA2, jet energies are measured using conventional calorimetry and the resolution so obtained for the two-jet invariant mass in the region of the IVB masses is $\approx 8$ GeV/c$^2$, a value far from adequate for the event sample available at present.

For these reasons, both experiments have chosen to detect the IVBs by identifying their leptonic decays:

\[
\begin{align*}
W^+ &\rightarrow e^+\nu_e (\bar{\nu}_e) \\
Z^0 &\rightarrow e^+e^- \\
W^- &\rightarrow \mu^-\nu_\mu (\bar{\nu}_\mu) \\
Z^0 &\rightarrow \mu^+\mu^- 
\end{align*}
\]

in both UA1 and UA2,

in UA1 only,

for which backgrounds are much lower, as we shall see later, in spite of the smaller branching ratios.

The decision to detect electrons instead of muons in the UA2 experiment was based on the possibility of measuring electron energies with good precision using compact electromagnetic calorimeters (typically only $\approx 25$ cm thick), whereas muon detection implies large magnetic field volumes to measure its momentum and the use of thick iron absorbers.
When searching for the decay $Z^0 \rightarrow e^+ e^-$, or $Z^0 \rightarrow \mu^+ \mu^-$, the $Z^0$ mass can be directly determined by measuring the energies (or momenta) of the two leptons and the angle $\alpha$ between their directions:

$$M^2 = 4E_1E_2 \sin^2 \frac{\alpha}{2} \quad (33)$$

and a good mass resolution $\Delta M$ (smaller than the natural width $\Gamma_z$) is essential to determine $\Gamma_z$.

The mass resolution $\Delta M$ can be estimated from a knowledge of the measuring errors $\Delta E_1$, $\Delta E_2$, and $\Delta \alpha$:

$$\frac{\Delta M}{M} = \frac{1}{2} \sqrt{\left(\frac{\Delta E_1}{E_1}\right)^2 + \left(\frac{\Delta E_2}{E_2}\right)^2 + \left(\frac{\Delta \alpha}{\tan \frac{\alpha}{2}}\right)^2} \quad (34)$$

The opening angle $\alpha$ is measured using tracking devices and its error $\Delta \alpha$ can easily be made as low as 0.01 rad, or less. Its effect on $\Delta M/M$ is then negligible, because at collider energies $\alpha$ is generally larger than 90°.

Electron energies are measured in calorimeters with a resolution $\Delta E/E = 0.15/\sqrt{E}$ ($E$ in GeV, see Section 3), giving

$$\frac{\Delta M}{M} \approx \frac{0.15}{\sqrt{M}} \quad (35)$$

where $M$ is in GeV/c$^2$, for a symmetric decay configuration. For $M \approx 90$ GeV/c$^2$ the mass resolution is $\Delta M \approx 1.4$ GeV/c$^2$, a value adequate to measure $\Gamma_z$.

For the decay $Z^0 \rightarrow \mu^+ \mu^-$ muon momenta $p_1$ and $p_2$ replace the electron energies in Eqs. (33) and (34). They are determined by measuring the muon deflection in the magnetic field with a resolution which is typically $\Delta p/p \approx 0.5\%$ $p$ (in GeV/c) in the UA1 detector (there is only a 0.6 mm sagitta in a 1 m long track from a 45 GeV/c particle). In this case we obtain

$$\frac{\Delta M}{M} \approx 1.8 \times 10^{-3} M \quad (36)$$

where $M$ is in GeV/c$^2$, which corresponds to a mass resolution $\Delta M \approx 14$ GeV/c$^2$ at $M \approx 90$ GeV/c$^2$, a value much worse than in the $e^+ e^-$ case. Furthermore, $\Delta M/M$ increases linearly with $M$, whereas for $e^+ e^-$ pairs it decreases as $M^{-\frac{1}{2}}$, showing the superiority of the $e^+ e^-$ channel in searching for possible new particles with masses larger than those of the IVBs.

In the case of the decay $Z^0 \rightarrow e^+ e^-$ the actual mass resolution is worse than that given by Eq. (35) because it is difficult to keep the calorimeter calibration under control over the long period of time during which data are collected. At the pp
Collider there is no heavy particle of known mass decaying into $e^+e^-$ pairs whose production rate is high enough to be used as a calibration standard. The calibration is monitored, therefore, by a variety of methods, such as the use of radioactive sources, light pulsers, etc., and by periodically recalibrating as many calorimeter modules as possible with electron beams of known energy.

The calibration uncertainty increases $\Delta M/M$ by a systematic contribution of the order of 1.5 to 3% to be added in quadrature to Eq. (35).

The detection of the decay mode $W^\pm \rightarrow \ell^\pm \nu(\bar{\nu})$, where $\ell$ is either an electron or a muon, is based on a completely different method, because the neutrino cannot be detected. In the $W$ rest frame the lepton energy is just $M_W/2$. If we neglect the $W$ transverse momentum $p_T^W$, the lepton transverse momentum $p_T^\ell = (M_W/2) \sin \theta^*$, where $\sin \theta^*$ is the decay angle with respect to the beam direction in the $W$ rest frame, is a Lorentz invariant, and the $p_T^\ell$ spectrum can be obtained from the decay angular distribution $f(\theta^*)$ by a simple change of variables:

$$\frac{dn}{dp_T^\ell} \propto \frac{1}{\sqrt{1 - (2p_T^\ell/M_W)^2}} f\left[\sin^{-1}\left(\frac{2p_T^\ell}{M_W}\right)\right].$$

The singularity at $p_T^\ell = M_W/2$ arises only from the change of variables and it is often called the Jacobian peak. The $W$ transverse motion and the $W$ natural width have the effect of smearing this peak. However, if the mean value of $p_T^\ell$ is much smaller than $M_W/2$, the Jacobian peak is still the most distinctive feature of the charged-lepton $p_T$ distribution in the decay $W \rightarrow \ell\nu$.

Experimentally, the problem consists in detecting high-$p_T$ charged leptons (either $e^-$ or $\mu^-$) over the largest possible solid angle, and in searching for a peak structure in their $p_T$ spectrum. A fit to such a structure using the form expected for the Jacobian peak from $W \rightarrow \ell\nu$ is a method of determining the $W$ mass.

5.6 Background to the lepton signal

In a search for high-$p_T$ leptons the main background is represented by high-$p_T$ hadrons, or jets of hadrons, which are misidentified as leptons. To illustrate the relative level of the signal and of the main background source, in Fig. 28 the inclusive $p_T$ spectrum of hadronic jets is compared with that expected from $W \rightarrow \ell\nu$ decay. Also shown in Fig. 28 is the inclusive spectrum of high-$p_T$ single hadrons, which can be obtained from the jet $p_T$ spectrum using the jet fragmentation function $D(z)$ (see Section 4.6):

$$\frac{d\sigma}{dp_T^\ell} = \int_{p_T^\ell}^{p_T^\text{max}} \frac{d\sigma}{p_T^\ell} \frac{d\sigma}{dp_T^\text{jet}} D(p_T^\ell/p_T^\text{jet}).$$
Fig. 28 Comparison of the inclusive jet yield at the Collider with the yield of single high-$p_T$ hadrons and of high-$p_T$ leptons expected from the decay of the IVBs.

We see that, near the Jacobian peak, the relative rates are approximately in the ratios jets:single hadrons:leptons $\approx 10^3:1:1$. This is a very favourable experimental situation because rejection factors against multi-hadrons of the order of $10^4$, or larger, are not difficult to achieve.

An additional method of rejecting backgrounds when searching for the decay $W \rightarrow \ell\nu$ consists in detecting the transverse-momentum imbalance resulting from the fact that the event contains a high-$p_T$ neutrino which escapes detection. We write

$$\vec{p}_T = \vec{p}_W - \vec{p}_T \tag{39}$$

where $\vec{p}_T$ is the transverse momentum of the detected lepton and $\vec{p}_W$ is equal and opposite to the vector sum of the transverse momenta of all other particles observed in the same event [Eq. (39) cannot be extended to the longitudinal components because the energy carried by undetected particles produced at very small angles to the beams is generally too large to be neglected]. For small values of $p_T$, as is often the case, one expects $\vec{p}_T \approx -\vec{p}_T$.

On the other hand, most background events result from misidentified high-$p_T$ hadrons, which, as discussed in detail in Section 4, belong in general to events consisting of two jets with approximately equal and opposite transverse momenta. In this case the main contribution to $\vec{p}_T$ is represented by the jet opposite to the mis-identified high-$p_T$ lepton, and one finds $\vec{p}_T \approx 0$. 

121
This method is not applicable, of course, to the decay $\pi^0 \rightarrow \mu^+ \mu^-$. However, in this case the two-jet background is reduced by the square of the rejection factor against hadrons, because the experiment involves the simultaneous detection of two charged leptons.

5.7 Lepton detection in the UA experiments

High-$p_T$ electrons are expected to develop electromagnetic showers which are almost fully contained in the electromagnetic compartments of the UA1 and UA2 experiments, with only a small fraction of the energy leaking into the hadronic compartments behind them. In addition, the lateral size of the shower must be small enough to be compatible with that expected for electrons (at most two adjacent cells in UA1 and a matrix of $2 \times 2$ cells in the UA2 calorimeter which has a finer segmentation).

However, collimated jets whose fragmentation is dominated by $\pi^0$ mesons decaying into photons satisfy these requirements. To reject this background additional criteria must be used.

In the UA1 experiment the tracking detector in the large magnetic field volume is used to require the presence of a high-momentum charged particle which points to the energy cluster in the calorimeter. In the case of genuine electrons the momentum $p$ and the energy $E$, as measured in the calorimeter, must agree, and the condition $|p^{-1} - E^{-1}| < 3\sigma$ is imposed, where $\sigma$ is the r.m.s. error on the difference $p^{-1} - E^{-1}$ (the main contribution to $\sigma$ results from the measuring error on $p$).

In the UA2 experiment, this condition is only applied in the two forward regions, where a magnetic field is available. In the central region, $40^\circ < \theta < 140^\circ$, no momentum measurement is possible and the presence of a charged-particle track pointing to the energy cluster is not enough, in general, to reject high-$p_T$ jets in which most of the energy is carried by $\pi^0$ mesons. If the observed track belongs to an electron, the electromagnetic shower will begin in most cases in the $\sim 1.5$ radiation lengths thick preshower counter located in front of the calorimeter (see Section 3.3), where both the pulse height of the signal and its position in space can be measured. In this case the distance between the preshower signal and the track is consistent with the resolution (typically 3 mm in the central detector and $\sim 10$ mm in the forward ones).

On the contrary, if the track belongs to a charged hadron and the calorimeter energy deposition is due to photons, large signals in the preshower counter, if any, will be observed at larger distances from the track.

In order to further reduce the background from high-$p_T$ jets, both experiments require the absence of additional large-energy clusters in the calorimeter cells adjacent to those hit by the electron, and, in the regions where a magnetic field exists, the absence of other high-momentum tracks nearby.
High-$p_T$ muons are detected in UA1 by requiring that a charged-particle track is observed in the muon chambers which surround the detector. To reach these chambers, particles must traverse all calorimeters and additional iron absorber located after the magnet yoke (a total of $\sim$10 absorption lengths) without interacting. The muon track is required to match a high-momentum track measured in the central detector, within the resolution.

An independent determination of the muon momentum, with a precision $\Delta p/p$ of the order of 20%, is performed by measuring the track deflection in the magnet yoke.

As already mentioned in Section 3.3, no muon detection is attempted in the UA2 experiment.

5.8 Observation of the decay $W \rightarrow e\nu$

The first experimental evidence for the process $p + \bar{p} \rightarrow W^+ +$ anything, followed by its decay into an electron and a neutrino, was presented in January 1983, soon after the end of the 1982 physics run. In a data sample corresponding to a total integrated luminosity of about 20 nb$^{-1}$, the UA1 experiment observed six events containing an electron with $p_T > 15$ GeV/c as the only visible high-$p_T$ particle$^83)$. At the same time, four events with very similar characteristics were observed in the UA2 experiment$^84)$. Figure 29 shows the $p_T$ distribution of these ten electrons together. A hint of the Jacobian peak is already visible in this distribution, and a preliminary value of the $W$ mass ($M_W = 81 \pm 5$ GeV/c$^2$) was extracted from these data.

The total integrated luminosity collected by each experiment at the end of the physics run which took place during the spring of 1983 amounts to $\sim 130$ nb$^{-1}$. In the UA1 experiment$^85$), all events containing an electron candidate with $p_T > 15$ GeV/c

Fig. 29 First observation of the decay $W \rightarrow e\nu$: the electron transverse-momentum distribution in events with missing transverse energy obtained by the UA1 and UA2 experiments at the end of the 1982 running period.
and a hadronic jet within an azimuthal interval of $\pm 30^\circ$ opposite to the electron direction are rejected (a total of 291 events) from an initial sample of 343 events. A further fiducial cut is applied to the remaining 52 events to exclude electrons which hit the large-angle electromagnetic calorimeter (see Section 3.2) within $\pm 15^\circ$ of their top and bottom edges. A total of 43 events survive this cut, with the $p_T$ distribution as shown in Fig. 30.

In these events, the transverse momentum of the undetected neutrino, $p_T^\nu$, is obtained as discussed in Section 5.6. Figure 31 shows the value of the $p_T^\nu$ component opposite to the electron direction, defined as $p_{\text{opp}}^\nu = -p_T^\nu \cdot p_T^e / p_T$, as a function of the electron transverse momentum $p_T^e$. As expected for $W$ production with $p_T^W \ll m_W$, in all of the events the neutrino transverse momentum is approximately equal and opposite to that of the electron.

A value of the $W$ mass is extracted from this event sample by defining for each event a transverse mass variable:

$$M_T^2 = 2 p_T^e p_T^\nu (1 - \cos \Delta \phi), \quad (40)$$

where $\Delta \phi$ is the azimuthal separation between $p_T^e$ and $p_T^\nu$. The advantage of using $M_T$ is that one can obtain theoretical $M_T$ distributions with no need to know the $W$ transverse motion. Fitting the experimental $M_T$ distribution to the expected one, where $M_W$ is considered as a free parameter, the UA1 Collaboration quote the value

$$M_W = 80.9 \pm 1.5 \text{ (stat.)} \pm 2.4 \text{ (syst.) GeV/c}^2, \quad (41)$$

where the systematic error reflects the $\pm 3\%$ uncertainty on the energy scale of the calorimeters.

Fig. 30 The electron transverse-momentum distribution for the whole data sample collected in the UA1 experiment.
The UA2 results\textsuperscript{86}) are based on a sample of 134 events which contain an electron candidate with $p_T > 15$ GeV/c. This sample contains the events in which hadronic jets appear at opposite azimuthal angles to the electron. These events have been rejected from the final UA1 sample. Figure 32 shows the position of these events in the $p_T$, $\not{p}_T$ plane. Two regions of relatively high population density are clearly visible in this plot, one in the region where $\not{p}_T \approx p_T^e$, as expected from $W \rightarrow e\nu$ decay, and the other near $\not{p}_T \approx 0$. In the events belonging to this latter region of the plot, the electron transverse momentum is balanced by another high-$p_T$ particle or multi-particle system. As we shall see in Section 5.10, the events with very high values of $p_T^e$ ($p_T^e > 25$ GeV/c) and $\not{p}_T \approx 0$ are mostly $Z^0 \rightarrow e^+e^-$ decays.

The 35 events with $\not{p}_T > 0.8 p_T^e$ are considered as $W \rightarrow e\nu$ candidates. Their $p_T^e$ distribution is shown in Fig. 33a. A fit to the $W$ mass from a recent publication\textsuperscript{87}) gives

$$m_W = 83.1 \pm 1.9 \text{ (stat.)} \pm 1.3 \text{ (syst.) GeV/c}^2,$$

where the systematic error reflects the uncertainty on the energy scale of the UA2 electromagnetic calorimeters. The best fit curve to the experimental $p_T^e$ distribution, also shown in Fig. 33a, is obtained as the sum of four contributions: i) $W \rightarrow e\nu$ decays, which dominate for $p_T^e > 25$ GeV/c; ii) electrons from the decay chain $W \rightarrow \tau\nu$, $\tau \rightarrow e\nu\nu$; iii) $Z^0 \rightarrow e^+e^-$ decays with one electron falling outside the UA2 acceptance; and iv) a background of fake electrons from misidentified hadrons, which dominate near threshold. These contributions are shown separately in Fig. 33b.
Fig. 32  Correlation between the neutrino and electron transverse momenta in the UA2 experiment.

Fig. 33  a) Electron transverse-momentum distribution for events with missing transverse energy in the UA2 experiment. The solid line is the result of a best fit as explained in the text; b) the contributions from $W \rightarrow e\nu$, $W \rightarrow \tau\nu$ followed by $\tau \rightarrow e\nu\bar{\nu}$, $Z^0 \rightarrow e^+e^-$ with one electron undetected, and background are shown separately.
Finally, in Fig. 34 we show the $W$ transverse momentum distribution, as measured in UA1, together with a number of QCD predictions. The events having the highest values of $p_T^{W}$ shown as dark areas in Fig. 34, contain a clearly visible hadronic jet which balances the $W$ transverse momentum.

5.9 Charge asymmetry in $W \rightarrow e \nu$ decay

In a theory of the weak interactions with $V-A$ coupling, it is well known that only fermions in a state of helicity $-1$ (spin antiparallel to the momentum) and anti-fermions with helicity $+1$ can couple to the $W$ field. As a consequence, in the elementary processes, $u + d \rightarrow W^{+}$ and $d + u \rightarrow W^{-}$, the $W^{\pm}$ is fully polarized along the direction of the incident $p$, because the fermions are supplied by the incident $p$ and the antifermions by the $\bar{p}$.

By applying the same arguments to the outgoing fermion-antifermion pairs ($e^{-}\bar{\nu}_{e}$ or $e^{+}\nu_{e}$), it can easily be shown that $e^{-}$ emission along the $\bar{p}$ direction, and $e^{+}$ emission along the $p$ direction, are forbidden by angular momentum conservation. Quantitatively, if $\Theta^{A}$ is the angle between the charged lepton and the proton direction in the $W$ rest frame, the angular distribution is given by

$$\frac{dn}{d \cos \Theta^{A}} = \frac{1}{2} \left( 1 - \epsilon \cos \Theta^{A} \right)^2,$$

where $\epsilon = -1$ for electrons and $+1$ for positrons.
5.10 Observation of the $Z^0$ boson

Figure 37 summarizes the search for $Z^0 \rightarrow e^+e^-$ events in the UA1 experiment$^{93}$.

A first selection, based on the presence of two energy clusters with $p_T > 25$ GeV/c in the electromagnetic calorimeter, produces a sample of 152 events. Second-level cuts requiring that for at least one of the clusters there is an isolated track pointing to it, and no energy deposition larger than 0.8 GeV is present in the hadronic calorimeter behind the cluster, reduce the sample to six events, with an invariant mass distribution which shows already an accumulation of four events in the region of the expected $Z^0$ mass value. These four events are the only ones which survive the application of the second-level cuts to both energy clusters.

Figure 38a displays all vertex-associated tracks and calorimeter hits in one of the four events. If only tracks and energy depositions having $p_T > 2$ GeV/c are displayed, the only visible tracks and energy clusters are those associated with the $e^+e^-$ pair (Fig. 38b).

The results of a search for the decay $Z^0 \rightarrow \mu^+\mu^-$ in the UA1 experiment have also been reported for a data sample corresponding to $\sim 40\%$ of the entire data sample. Out of 42 events containing a muon-candidate track with $p_T > 7$ GeV/c, only one event

![Figure 37](image-url)

**Fig. 37** Search for $Z^0 \rightarrow e^+e^-$ in the UA1 experiment:

a) first selection based on calorimetry only;

b) a high-$p_T$ isolated track points to at least one of the energy clusters;

c) both clusters have a high-$p_T$ isolated track pointing to them.
Fig. 38  a) Display of a $Z^0 \rightarrow e^+e^-$ event. All tracks and calorimeter hits are shown;  
b) the same event, where only tracks and calorimeter hits having $p_T > 2$ GeV/c are shown.
contains two muons. This event is shown in Figs. 39a and 39b. In this case one sees a clear hadronic jet with $p_T \approx 24$ GeV/c in addition to the two high-$p_T$ muons. Under the assumption that this jet balances the $p_T$-value of the muon pair, the dimuon invariant mass can be determined from the knowledge of the muon directions and the jet.

![Event 6600. 222.](image1)

**Fig. 39** a) Display of the $Z^0 \rightarrow \mu^+\mu^-$ event. All tracks and calorimeter hits are shown;

b) the same event, where only tracks and calorimeter hits having $p_T > 2$ GeV/c are shown. An additional high-$p_T$ jet is clearly visible.
transverse momentum vector, which is measured in the calorimeter. The result is $M_{\mu\mu} = 95.5 \pm 7.3 \text{ GeV}/c^2$, a value much more precise than that derived using only the muon momenta ($M_{\mu\mu} = 113^{+25}_{-16} \text{ GeV}/c^2$).

A summary of the mass values of the five events together is shown in Fig. 40. The value

$$M_Z = 95.2 \pm 1.4 \text{ (stat.)} \pm 2.9 \text{ (syst.) GeV}/c^2,$$  \hspace{1cm} (44)

is obtained for the $Z^0$ mass\(^{91}\), where the systematic error reflects the uncertainty on the energy scale of the calorimeters.

A fit of the mass distribution of the four $Z^0 \rightarrow e^+e^-$ events to a Breit-Wigner shape gives $\Gamma_Z < 8.5 \text{ GeV}/c^2$ at the 90% confidence level\(^{91}\). As discussed in Section 5.3, this upper limit can be translated into an upper limit for the number of neutrino types with $m_\nu < M_Z/2$ within the standard model. The result is $N_\nu \leq 35$, a number much lower than previous upper limits from particle-physics experiments.

The UA2 results\(^ {92}\) from a search for $Z^0 \rightarrow e^+e^-$ events are illustrated in Fig. 41. A first event selection based on calorimetric information only (consistency of energy cluster size and leakage into the hadronic compartment with the expectations from an electron shower) produces a sample of 24 events containing two energy clusters with an invariant mass above a threshold of 50 GeV/$c^2$ (Fig. 41a). When the electron-identification criteria described in Section 5.7 are imposed on at least one of the two clusters, only eight events survive, all falling in a narrow mass interval near the expected $Z^0$ mass value (Fig. 41b). The invariant mass of the $e^+e^-$ pair can be determined unambiguously only in four events [we refer the reader to the original UA2 paper\(^ {92}\) for a discussion of the remaining four events]. The result is

$$M_Z = 92.7 \pm 1.7 \text{ (stat.)} \pm 1.4 \text{ (syst.) GeV}/c^2,$$  \hspace{1cm} (45)

![Fig. 40 Invariant masses of lepton pairs in the UA1 experiment.](image)
Fig. 41  Search for $Z^0 \rightarrow e^+e^-$ in the UA2 experiment:

a) First event selection based on calorimetry only;

b) at least one of the two energy clusters satisfies the electron-identification criteria.

where the systematic error reflects the uncertainty in the energy scale of the calorimeter. The r.m.s. deviation of these four events from the value (45) is 2.0 GeV/c², which is consistent with the experimental mass resolution and gives an upper limit $\Gamma_Z \leq 6.5$ GeV/c² for the $Z^0$ mass.

One of the UA2 events is identified as a $Z^0 \rightarrow e^+e^-\gamma$ decay, with a 24 GeV photon and 11 GeV electron clearly separated by an angle of 31°. Figure 42 shows the transverse energy distribution over the calorimeter cells for this event, and its transverse view. The probability that a $Z^0 \rightarrow e^+e^-$ decay appears in such a configuration or in less probable ones as a result of radiative corrections is of the order of 1%, so it is not so surprising to find one such event in a sample of eight. However, a second $Z^0 \rightarrow e^+e^-\gamma$ event and a $Z^0 \rightarrow \mu^+\mu^-\gamma$ event have recently been reported by the UA1 Collaboration, with probabilities of the order of 1% if only radiative corrections are considered (the $Z^0 \rightarrow e^+e^-\gamma$ event in the UA1 experiment is actually one of the four original $Z^0 \rightarrow e^+e^-$ events, which is interpreted to be a $Z^0 \rightarrow e^+e^-\gamma$ decay after a recalibration of the calorimeter). These observations open up the possibility of a new and unexpected decay mode of the $Z^0$ boson.
Fig. 42 The $Z^0 \rightarrow e^+ e^- \gamma$ event observed in the UA2 experiment:

a) transverse-energy distribution over the calorimeter cells;

b) transverse view of the event. The dotted line joins the event vertex to the space position of the signal detected in the wire chamber following the tungsten converter. It represents, therefore, the line of flight of the photon.
5.11 Comparison with the standard-model predictions

The values of the $W$ mass obtained in the UA1 and UA2 experiments [Eqs. (41) and (42), respectively], can be used together with Eq. (11) to derive a value for $\sin^2 \theta_W$. If we average the two results after adding in quadrature their respective statistical and systematic errors we obtain

$$\sin^2 \theta_W = 0.221 \pm 0.010$$

in good agreement with low-energy experiments\(^{70}\).

The measurement of the ratio $M_Z/M_W$ has the advantage that most of the systematic uncertainties in the mass scale of the experiments cancel out. If we rewrite Eq. (17) as

$$\rho = \frac{M_Z^2}{M_W^2 \cos^2 \theta_W},$$

we obtain $\rho = 0.926 \pm 0.041$ from the UA1 experiment\(^9^{11}\), and $\rho = 1.03 \pm 0.06$ from the UA2 experiment\(^8^{77}\). The weighted average of the two results is

$$\rho = 0.959 \pm 0.034$$

in good agreement with the predictions of the standard model\(^6^{95}\).

6. CONCLUSIONS

The first $p\bar{p}$ collisions at $\sqrt{s} = 540$ GeV were observed in July 1981\(^{135}\). In less than two years since that date, the CERN $p\bar{p}$ Collider has already achieved all of the goals for which it had been proposed. Both the charged and neutral IVBs have been discovered, and all their measured properties have been found to be in good agreement with the predictions of the standard model. It is remarkable that not only the magnitude of the signal was correctly predicted theoretically, but also the background, which consists mainly of high-$p_T$ hadronic jets, turned out to be at the level predicted by QCD. In this respect, the discovery of the IVBs was an easy experimental task.

For the first time after the discovery of high-$p_T$ phenomena more than ten years ago, high-$p_T$ hadronic jets have been unambiguously seen in experiments without trigger bias, and it has become possible to study hard parton scattering in very favourable experimental conditions. Here too all the results obtained so far are in good agreement with theoretical expectations.
To conclude, the results achieved so far at the CERN pp Collider are mostly to the merit of those who proposed to build this facility, those who built it, and those who predicted the phenomena to be found there. However, we must remember that every newly opened energy domain has uncovered unanticipated discoveries. Let us hope that this will be the case for the CERN pp Collider as well.

References and Footnotes

The energy loss suffered by an electron of energy $E$ on a circular orbit of radius $R$ is $\Delta E = 0.0885E^4/R$ MeV/revolution ($E$ in GeV, $R$ in metres).
M. Glück et al., Z. Phys. 1982, C13, p. 119.
61. TASSO Collaboration, private communication to NA1.
73. No evidence for the production of t̅e pairs has been found at the highest e+e− collision energy available at PETRA, √s = 45.2 GeV.
79. F.E. Paige, Updated estimates of W production in pp and p̅p interactions,
86. The UA2 Collaboration, Latest results from the UA2 experiment at the CERN p̅p
   Collider, Proc. Int. Europhysics Conf. on High Energy Physics, Brighton,
   1983 (eds. J. Guy and C. Costain) (Rutherford Appleton Lab., Didcot, 1983),
   p. 472.
87. P. Bagnaia et al., A study of high transverse momentum electrons produced in
   Z. Phys. C.
   See also ref. 82.
91. The UA1 Collaboration, Experimental observation of lepton pairs of invariant
   mass around 95 GeV/c² at the CERN p̅p Collider, Proc. 1983 Int. Symposium on
   Lepton and Photon Interactions at High Energies, Cornell University, 1983.
93. F.A. Berends and R. Keiss, Hard photon effects in W⁺ and Z⁰ decay, University
   of Leiden, 1983.
95. The staff of the CERN proton-antiproton project, Phys. Lett. 1981, 107B,
   p. 306.
96. T.D. Lee, quoted by L.M. Lederman in Proc. of the Third Topical Workshop on
   p. 577.